Mechanisms of transcriptional regulation by SRF co-factors

Francesco Gualdrini

University College London

and

Cancer Research UK London Research Institute

PhD Supervisor: Richard Treisman

A thesis submitted for the degree of

Doctor of Philosophy

University College London

September 2014

Declaration

I, Francesco Gualdrini, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Part of the work presented was done in collaboration:

The bioinformatics analysis presented in Chapter 2, 3, 4 and 7 was done in collaboration with the Bioinformatics and Biostatistics group at LRI.

The preparations of the Chromatin and RNA libraries for the genomic studies were performed in collaboration with the Advanced Sequencing facility at LRI.

The investigation of the role of the TCFs in establishing chromatin changes in response to Ras/Erk activation was done in collaboration with Cyril Esnault. Figure 6.1 was entirely produced by Cyril Esnault.

Some of the data presented in this thesis have been used in a recent publication:

Cyril Esnault, Aengus Stewart, Francesco Gualdrini, Phil East, Stuart Horswell, Nik Matthews, and Richard Treisman (2014). Rho-actin signalling to the MRTF co-activators dominates the immediate transcriptional response to serum in fibroblasts. Genes Dev. 28: 943–958.

Abstract

Serum response factor (SRF) controls gene activation in response to changes in actin dynamics and mitogen-activated protein kinases. SRF has low intrinsic transcriptional activity and requires the recruitment of one of two families of co-activators: the MRTFs (myocardin-related transcription factors) and the TCFs (ternary complex factors). MRTFs are actin-binding proteins. Disruption of the actin-MRTF interaction is sufficient to induce MRTF nuclear accumulation and transcriptional activation. The TCF family are specifically activated by MAPK signalling. This thesis aims to elucidate how the SRF transcription network is controlled. The work presented encompasses two projects focused on each of the co-activator families.

The regulation of MRTF shuttling from the cytoplasm to the nucleus is relatively well understood while its regulation once in the nucleus is still uncharacterized. The work demonstrates that nuclear MRTF activities are influenced by nuclear actin. Nuclear actin interferes with MRTF-DNA binding and target gene activation. In the presence of G-actin, nuclear MRTF can associate with target loci and recruit Pol II that, although traverses the gene body, does not generate stable mRNA. This inhibited state is accompanied by hypo-phosphorylation of the Pol II CTD. Dissociation of MRTF-actin interaction is required to re-establish Pol II phosphorylation and mRNA accumulation.

The Erk-TCF signalling pathway was used to investigate how chromatin signatures are established in response to cues. Fibroblasts lacking all three TCFs, or reconstituted with mutant derivatives of the Elk-1 TCF were generated. Following Erk activation, chromatin immunoprecipitation and RNA-sequencing techniques, were employed to study the role of the TCFs in chromatin changes and transcriptional activation. It was possible to show that signal-induced chromatin changes occur in absence of transcription, and the specific chromatin signature requires Elk-1 DNA binding and phosphorylation. In addition analysis of the H3K27me3 mark demonstrated that Elk-1 activation is required to maintain a permissive chromatin landscape.

Acknowledgement

I thank Richard for the opportunity to work in his lab, for all the tough discussions, support and for having challenged me with inspiring projects.

I thank Cyril, it was great to work with him. For the mentoring, the shouting, the fun at conferences and the trust that made our collaboration enormously fruitful.

I thank everyone in the lab for all the support and discussions. Thanks to Rob for his help and pub chats, to Patrick who always kept a sense of humour even when everyone lost theirs. Thanks to the French speaking corner Diane and Jessica (and Cyril), and to Ricky for all the laughs and coffee breaks. A special thank goes to Matthew for the constant help, the calm and patient. I would like also to thank Lucy for her precious assistance.

I thank Sebastian Maurer for being a friend, for all the weekly climbs, the brainstorms at the pub and for having taught me how to purify proteins the 'German way'.

I thank Aengus Stewart, Phil East and Stuart Horswell and the Bioinformatics and Biostatistics group for the invaluable support. I also would like to thank Nik Matthews and all the people in the Advanced Sequencing facility for their expert assistance. Many thanks go to Svend Kjaer and Annabel Borg for all their hard work in the protein purification facility. A big thank you goes to Graham Clark and the Equipment Park for their tremendous work and assistance.

I thank Jesper Svejstrup and Julie Cooper for constructive advice and precious discussions and their optimism.

I thank my parents who never believed my Maths teacher at school and their support over the years. Thanks to my friends especially Cavaz, Zeno, Jack, Evi, Vincent, my goddaughter Lucie and my two brothers Michele and Giovanni for reminding me that enjoying life is good.

Most importantly, I am grateful to my wife, Silvia, who has been incessantly helpful and patient enough to stay with me no matter what.

Finally I thank Cancer Research UK and the European Research Council for enabling and funding this work.

Table of Contents

Table of figures

List of tables

Abbreviations

aa	Aminoacid
ABP	Actin binding protein
AD	Activation domain
ADF	actin-depolymerising factor ADP adenosine diphosphate
ARP	actin-related protein
bp	Base pair
CAGE	Cap analysis gene expression
CBP	CREB-binding protein
CD	Cytochalasin D
CE	Capping Enzyme
ChIP	Chromatin immunoprecipitation
CoTC	co-transcriptional cleavage
CR3	conserved region 3
CREB	cAMP response element-binding protein
CRM1	Chromosome Region Maintenance (Exportin 1)
CTD	Carboxyl-terminal domain
DBD	DNA binding domain
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
DRB	5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole
DSE	downstream sequence element
DSIF	DRB Sensitivity Inducing Factor
E1A	Adenovirus early region 1A
Elk-1	ETS-Like Gene 1
ERK	extracellular signal regulated kinase
Esc	embryonic stem cells
ESN	Elk-1-/-SAP-1-/-Net ∂ / ∂
ETS	E26 transformation-specific or E-twenty -six
F-actin	Filamentous actin
FACT	facilitates chromatin transcription

FCS	foetal calf serum
FOX	Forkhead box
FRET	Förster resonance energy transfer
G-actin	Globular (monomeric) actin
GAP	GTPase activating protein
GDI	guanine nucleotide-dissociation inhibitor
GEF	guanine nucleotide exchange factor
GTF	General transcription factor
H	Histone
HA	hemagglutinin
HAT	Histone acetyltransferases
HDAC	Histone deacetylase
HDE	Histone downstream element
HOX	Homeobox
HSV	Herpes simplex virus
IEG	Immediate earlygene
IP	immunoprecipitation
IP3	inositol trisphosphate
Kb	kilobases
KDM	Histone lysine demethylase
KMT	Histone lysine methytransferase
LatB	latrunculin B
LMB	leptomycin B
LZ	leucine zipper
MADS	MCM1, AG, DEFA, SRF
MAPK	mitogen-activated protein kinase
MC	Myocardin
MLL	mixed lineage leukemia
MRTF	Myocardin related transcription factor
NELF	Negative elongation factor
NLS	Nuclear localisation signal
NR	Nuclear receptor
OC	Open Complex
ORF	Open reading frame

p-TEFb	positive transcription elongation factor
PAF	polymerase associated factor
PAS	Polyadenylation site
PIC	Pre-initiation complex
PIP2	Phosphatidylinositol-4,5-bisphopshate
PKC	Protein kinase C
PLC	phospholipase C
PLK3	polo-like kinase 3
Pol	RNA polymerase
PTKR	Protein Tyrosine Kinase Receptor
PTM	Posttranslational modification
RAS	Rat sarcoma
RasGRP	Ras guanyl nucleotide releasing protein
RNA	ribonucleic acid
RNAi	RNA interference
RNGTT	RNA guanylyltransferase and 5'-phosphatase
RNMT	RNA (guanine-7-) methyltransferase
RPEL	motif name (RPxxxEL); PFAM 2755
Rpm	revolutions per minute
SAP	SAF-AIB, Acinus, Pias
SAP1	SRF accessory protein 1
SEC	super elongation complex
Ser	Serine
SET	Su(var)3-9, Enhancer of Zeste, Trithorax
shRNA	short hairpin RNA
snRNP	small nuclear ribonucleic particles
SP	Sreine-Proline
SR	Steroid receptor
SRE	Serum Response Element
SRF	Serum response factor
TAF	TBP associated factor
TBP	TATA-binding protein
TCF	Ternary Complex Factor
TES	transcription end sites

TF	Transcription factor
Thr	Threonine
TPA	12-O-Tetradecanoylphorbol-13-acetate
TSS	Transcription start site
Tyr	Tyrosine
UTR	untranslated region
VP16	Herpes simplex virus protein 16
WT	wild-type
α–	anti-(in ChIP)

Chapter 1. Introduction

1.1 Signalling to Transcription

Cells are capable of recognising and responding to particular stimuli outside or inside their cellular environment. This collection of complex signals is matched to sophisticated intracellular responses. The molecular co-evolution of signals and responses was the main route through which metazoan life evolved (Pires-daSilva & Sommer 2003).The connection between signal transduction and control of gene expression is the backbone of modern developmental genetics and cellular biochemistry.

Signal-dependent gene regulation operates throughout the life of cells, tissues, organs and organisms. All these processes require the production of messenger RNA from protein coding genes, which is carried out by the RNA polymerase II (Pol II). In synthesis, the elaboration of the recorded signals leads to the assembly of the initiation complex and Pol II recruitment that gives rise to the primary transcript that ultimately becomes mRNA (see Chapter 1.3).

Members of a large family of proteins, the Transcription Factors (TFs, see below in this chapter), coordinate signal-dependent gene activation. TFs, between 1000 and 2000 in humans and mice (H. M. Zhang et al. 2012b), display two characteristic domains: a DNA binding domain and a regulatory domain. TFs are site-specific factors able to recruit co-regulators and the transcriptional machinery producing a specific transcriptional program.

With the transcriptome currently estimated to include over 20000 protein coding genes and about 7000 non coding genes in mammals (Pruitt et al. 2013), what gives a unique spatial and temporal expression to each gene is the combinatorial use of transcription factors and co-regulators (Carninci et al. 2005).

1.2 Regulatory transcription factors

Research on organismal evolution, from unicellular organisms to complex systems, has shown that the degree of complexity does not directly correlate with the number of genes but to the complexity of the transcriptional programming (Gregory 2005).

Higher systems, such as eukaryotes, have evolved a more complex way of controlling their gene expression than prokaryotes (Levine & Tjian 2003). The arrangement of encoded DNA information, the way the DNA is packed into chromatin and the combination of different TFs is what allows complex organisms to alter their cell types and growth patterns in different ways (Gregory 2005; Carninci et al. 2005; Levine & Tjian 2003).

Transcription factors are modular proteins consisting of two major domains: (I) a sequence-specific DNA-binding domain (DBD), that allows TFs to bind cis-regulatory DNA sequences either at enhancers or at promoters; and (II) a regulatory domain that allows TFs to receive and respond to signals, interact with the transcriptional machinery and influence the outcome of the transcriptional process. These domains are most of the time separable within the TF protein. However the DNA-binding and regulatory domain can also be embedded within the same amino-acid sequence, for example in the serum response factor (SRF) in metazoan or GAL4 in yeast (Ling et al. 1998; Mizutani & Tanaka 2003).

TFs can activate or repress transcription and they do so using diverse mechanisms such as: Pol II holo-enzyme direct recruitment, block or stabilisation; modification of the chromatin template *via* recruitment of re-modellers and modifiers; and recruitment of co-regulators.

1.2.1 Transcription co-regulators

TFs are often found in association with other factors called co-regulators. Co-regulators, including chromatin modifiers and re-modellers, while also playing crucial roles in controlling genes activation, lack DNA-binding domains and therefore are not classified as transcription factors (Schaefer et al. 2011). Co-regulators are specifically recruited by TFs, thus providing additional regulatory domains and functions. Examples are the CREB binding protein (CBP), recruited by the cyclic AMP response element-binding protein (CREB) (Mayr et al. 2001), and the Myocardin-related transcription factors (MRTFs), co-regulators of the serum response factor (SRF) (D. Z. Wang et al. 2002).

A recent classification of transcription co-factors comes from Schaefer and colleagues who considered proteins involved in transcription, interacting with TFs but not interacting with the DNA (Schaefer et al. 2011). The in-silico screen, based

on other databases and the web, led to the generation of the transcription co-factors database (TcoFs-DB). The TcoF-DB includes 1365 TFs and 529 co-regulators, of which 374 are hypothetical. This classification is just a starting point as most of the data is based on protein annotation, experiments and data submission that could be inaccurate and incomplete. Nevertheless up to a tenth of the human proteome is involved in transcription regulation and it is possible to appreciate that transcription co-regulator activities are diverse including, among others, signal transmission, control of TF-DNA binding and chromatin modification (Schaefer et al. 2011).

1.2.2 DNA binding domains and the principle of DNA recognition

The defining feature of a TF is its sequence-specific DNA binding domain. TFs are able to selectively distinguish DNA elements out of a vast number of sequences, and they do so through their DNA binding domain (Jolma et al. 2013). Each TF associates with its partially degenerated consensus sequence (about 6-16 nucleotides long). As will be seen below, TF sequence specificity is essential but not sufficient to uniquely specify their genomic distribution.

The focus of this introduction is about transcription related proteins but it is necessary to specify that there are other examples of sequence-specific DNA recognition among proteins that mediate recombination, DNA cleavage, and other processes. The general principle of DNA recognition is common between TFs and all these factors but the structural details differ.

The nucleic acid substrate presented to DNA binding proteins is a relatively uniform polymer. Negative charged sugar-phosphates constitute the main backbone of the DNA and a core of paired bases expose their functional groups in the major and minor grooves (Garvie & Wolberger 2001). As each base is chemically different, every given DNA sequence has its own “signature” determined by the pattern of functional groups exposed into the DNA grooves (Garvie & Wolberger 2001; Rohs et al. 2010). It is through a series of favourable electrostatic and van der Waals interactions that TFs are able to interact specifically with the DNA. Furthermore the DNA-binding domain of all TFs together with the DNA builds up a network of contacts that include salt bridges with positive side chains and hydrogen bonds with uncharged chains (Luscombe et al. 2001). In addition to

chemically driven interaction the DNA platform is distorted by TFs through α helices, β sheets and various secondary structure units that protrude into the DNA grooves forming specific contacts (Garvie & Wolberger 2001). DNA kinking and bending is not only a cause of lower free energy structures but in many cases affects TF activity (Rohs et al. 2010).

TFs can be classified according to their DNA-binding domains. The TFClass database stratifies TFs into five levels: superclass, class, family, genus and molecular species (Wingender et al. 2013). In total it is possible to distinguish nine superclasses (see Table 1.1), whose definition is based on the overall topology of their DBDs and the way they interact with the target DNA. Each superclass is then divided into several classes. TFs in the same class share a similar structure of the DNA-binding motif. An examples is SRF that belong to the class of MADS box factors and to the α-helices exposed by β-structure superclass. Each class is then divided into families based on sequence and functional similarities. For example the MADS box class comprises two families of TFs: TFs that regulate differentiation and TFs that respond to signals. Ultimately different families are composed of different genus, representing the physical TF encoded by a given gene, and molecular species when one gene produces more than one isoform.

How do TFs recognise their specific binding sites *in vivo*? How do structurally related DBD exert their unique function *in vivo*? How do different TFs influence each other? With modern genetics the assignment of TFs to their target genes becomes fundamental for a comprehensive understanding of all physiological, developmental and environmental conditions.

Knowing exclusively the DNA binding activity of each TF and their diverse affinity to consensus sequences is not sufficient to answer all these questions. Not all occurrences of a TF consensus element are bound by a TF in a given cell, and the location of TF binding sites in orthologous genes exhibits rapid evolutionary divergence (Villar et al. 2014). Nevertheless the DNA sequence is the primary driver of different TF binding (Wilson et al. 2008). The addition of a human chromosome into mouse liver cells leads to identical TF binding profiles and gene expression of the human chromosome when compared to human hepatocytes (Wilson et al. 2008). Despite this clear evidence we cannot predict functional TF binding patterns based solely on the DNA sequence (Odom et al. 2007; Borneman et al. 2007). Deciphering the protein-DNA recognition mechanism is crucial to

understand and potentially predict the DNA-binding specificity of uncharacterised proteins (Chu et al. 2009).

Classical genetics, chromatin immunoprecipitation followed by sequencing (ChIP-seq) and RNA silencing techniques (RNAi and shRNA), have being applied to dissect several transcriptional networks (Esnault et al. 2014; Rahl et al. 2010; X. Chen et al. 2008). Even in these cases, due to the large number of TFs, cell lines, environmental and developmental states the use of these methods is not sufficient to exhaustively understand transcriptional regulation.

1.2.3 DNA accessibility

As mentioned earlier most TF activities rely on their ability to recognise specific DNA sequences. DNA-binding affinity has been measured extensively for several TFs *in vitro*, and high-throughput methods have allowed the quantitation of TF binding preferences (Geertz & Maerkl 2010). *In vivo* analysis of TF genomic distributions using ChIP-seq approaches shows that most of the known TFs bind only a few per-cent of their potential consensuses in the genome (Carr & Biggin 1999; Joseph et al. 2010; Kaplan et al. 2011). *In vivo* the binding-site affinity of any given TF needs to be considered in the context of the chromatin.

Many TFs show little, if any affinity for their consensus when in the context of nucleosomes, therefore TFs seeking access to specific genomic loci need to compete directly with nucleosomes (Adams & Workman 1995). The thermodynamic model of this competition involves a straightforward reaction influenced by the TFs' affinity for its consensus and the relative concentration of both the nucleosomes and the TF in question (Raveh-Sadka et al. 2009). Co-occupancy of multiple TFs at the same loci also influences this model. Although the basis for this functional cooperation remains poorly understood, functional networks in which TFs co-localise together could be caused by specific or unspecific interactions between them at regulatory sequences (Ravasi et al. 2010; Voss et al. 2011). Consistent with this hypothesis DNAseI hypersensitive sites could be a result of multiple independent TF binding events as originally showed for the β-globin gene (Boyes & Felsenfeld 1996). Furthermore TFs clusters correlate with DNAseI hypersensitive sites across the genome (Thurman et al. 2012).

Nucleosome-TF competition has been invoked to explain why the majority of SRF binding sites become occupied only upon activation of the MRTF co-factors (Esnault et al. 2014). SRF DNA binding across the genome can be divided into constitutive and inducible sites. The former are embedded in broad regions of conserved motifs and nucleosome desert loci, consistent with the DNAseI hypersensitive maps (Thurman et al. 2012). The latter instead are associated with narrow regions of conservation occupied by nucleosomes under resting condition, suggesting that cooperative binding with other factors are less abundant and that it is the MRTF co-regulator recruitment that favours nucleosome displacement.

Nucleosomes also show different affinities for different DNA sequences (Sekinger et al. 2005; J. D. Anderson & Widom 2001). Different DNA sequences have diverse abilities to bend sharply, influencing the capability of the histone octamer to assemble with the DNA. Genome-wide studies have shown that it is possible to partially predict the disposition of nucleosomes across the genome on the basis of the DNA sequence (Segal et al. 2006). The combination of the TF binding and their cellular concentration with the affinity of the nucleosome to DNA sequences across the genome would potentially provide a tool to predict which consensus binding motifs are occupied in a given cell (Raveh-Sadka et al. 2009).

The access of TFs to highly condensed chromatin has been shown to require special features characteristic of a defined group of TFs called “pioneers” (Cirillo et al. 2002). Pioneer factors are chromatin binding factors able to recognise their target sequence in a nucleosomal context (for a review see Zaret & Carroll 2011). Studies on hepatogenesis lead to the discovery of FoxA and GATA, the first factors able to bind the condensed chromatin of the albumin gene enhancer in liver precursor cells (Cirillo et al. 2002). Members of the forkhead box protein A (FoxA) and FOXO families contain a C-terminal histone-binding domain required to loosen the condensed chromatin *in vivo* and *in vitro* (Zaret & Carroll 2011). Importantly the nucleosome de-compaction activity of FoxA does not require ATP or ATP-dependent chromatin remodellers (Cirillo et al. 2002). Furthermore the crystal structure of FoxA DBD showed a high similarity with the linker histone H1 and an avid nucleosome binding affinity (K. L. Clark et al. 1993). Although *in vitro* competition between FoxA and the histone H1 has not been demonstrated, *in vitro* biochemical studies extensively characterised the ability of FoxA to bind nucleosomes similarly, although less tightly, to histone H1 (Cirillo et al. 1998).

A different example comes from the glucocorticoid receptor (GR), which was found to bind its consensus specifically on nucleosomes (Perlmann & Wrange 1988). Recently it has been shown that GR works as an "assistant" helping factors to bind *via* the recruitment of ATP-dependent remodelling complexes (Voss et al. 2011).

A further example of this interdependence between TFs was recently shown for the "master regulators" Oct4, Sox2, Klf4, and c-Myc (Soufi et al. 2012). Oct4, Sox2 and Klf4 could all be defined as pioneer factors as they are all able to bind condensed chromatin allowing c-Myc to associate with this DNaseI-resistant landscape.

In conclusion each TF has its own strategy to reach its target sequence including: competition or specific affinity for nucleosomes, recruitment of chromatin-remodelling complexes and co-operative binding between multiple TFs.

1.2.4 Regulatory domains

If TFs' DNA-binding domains account for their correct positioning in the genome, TFs' activation (AD) or repression domains are responsible for their capacity to sense signals and mediate most of TF-protein interactions (Uesugi et al. 1997; Uesugi & Verdine 1999). It is through their activation domains that TFs are able to activate transcription, interacting with an elaborate set of general transcription factors (GTFs, see Chapter 1.3), co-regulators, chromatin remodellers, modifying enzymes and Pol II (see Chapter 1.3 and 1.4).

It has always been a challenge to efficiently classify ADs due to their unstructured nature (J. Liu et al. 2006). As ADs' activation properties cannot be easily recognised by sequence homology, several studies have tried to categorise them on the basis of polypeptide sequence (Mitchell & Tjian 1989): acidic activators like GAL4, GCN5 and VP16 (Sadowski et al. 1988), Glutamine-rich activators such as Sp1 (Courey & Tjian 1988) and proline-rich activators that include CTF and NF-1 (Mermod et al. 1989). Furthermore different features have been described within activation domains including hydrophobic and hydrophilic patches (Regier et al. 1993), amphipathic α-helices (Giniger & Ptashne 1987) and serine/threonine-rich patches (Cress & Triezenberg 1991; Drysdale et al. 1995).

Defining activation domains on the basis of their amino acid composition is somehow obsolete in the field as it is not clear if they truly reflect their activity. With advances in understanding transcription at the biochemical level, the classification of activation domains by structure, as opposed to sequence, has become more important. Interesting observations come from functional-structural studies of specific transactivation domains. In order to understand how TFs' regulatory domains work, the detailed description of the cocrystal or NMR structures of the activation domain together with their target is necessary. One of the most representative examples has been the herpes virion protein 16 (VP16) that is recruited to cellular targets through its interaction with TFs (Triezenberg et al. 1988). The carboxy-terminal part of VP16 (410-490aa) comprises two potent transactivation domains able to target several components of the basal RNA polymerase II machinery including TFIIA, TFIIB, TFIID, subunits of TFIIH and the Mediator complex (Kobayashi et al. 1995; D. B. Hall & Struhl 2002; Xiao et al. 1994; Uesugi et al. 1997; Mittler et al. 2003). Interaction between VP16's activation domain and target proteins is mediated by several acidic residues complemented by positive charges on the target protein (Jonker et al. 2005). These long-range electrostatic forces are responsible for an unstable complex. In this respect, hydrophobic residues are the key to obtain a stable interaction as side chains of these residues make important hydrophobic contacts with the target (Walker et al. 1993; Cress & Triezenberg 1991; Regier et al. 1993). Based on these observations, TADs are flexible domains in their free state able to form a structure upon interaction with their target proteins. Studies of the interaction between VP16 and Med25 have extended this view, proposing that VP16's AD is able to adapt to unrelated target surfaces (Vojnic et al. 2011). Similar concepts arise from studies based on c-Myc transactivation domain and its interaction with TBP. c-Myc interacts with TBP in two steps so that the folding of its TAD is coupled to the interaction with the TBP's surface (Hermann et al. 2001). Folding coupled to target binding could explain how TADs are able to interact with structurally different surfaces (Ferreira et al. 2005).

Other activation domains have been described using NMR studies. Clear examples come from the nuclear receptor with SRC-1, NcoA-1/STAT6, and p53-MDM2. A common feature of these interactions is a ΦXXΦΦ motif, where Φ is a hydrophobic residue within the transactivation domain. This motif can be the

archetypal LXXLL, common in nuclear receptors (Darimont et al. 1998), the FXXWL motif in p53 activation domain (Kussie et al. 1996) or domains within the Adenovirus early region 1A (E1a), able to interact with target proteins by forming amphipathic helices that plug into hydrophobic grooves (Pelka et al. 2008).

Conformational changes in activation domains can also be induced by post-translational modifications. In this case, the activation domain works as a switch able to record signals, adopting a permissive structure that can accommodate target surfaces. Section 1.5 will describe in detail the studies on the Elk-1 transactivation domain and its interaction with subunits of the Mediator complex.

1.2.5 Transcription factor regulation

Cells employ diverse mechanisms to efficiently control the accurate expression of genes required for cellular processes. Development, stem cell maintenance, environmental adaptation, cell cycle control and pathogenesis are only a few examples where TFs are causal effectors. Control of TF activity is therefore fundamental for accurate gene expression.

It is possible to distinguish mechanisms of regulation that directly affect the biochemical properties of TFs and mechanisms that affect TFs' functional behaviours. The former refers to the mode by which TFs are regulated and could include TFs' post-translational modifications (PTMs), synthesis and protein stability. The latter refers to general mechanisms employing a combination of mode of regulation that can affect, among others, TFs dimerization and oligomerisation, subcellular localisation and protein-protein interaction.

It is challenging to precisely separate each mechanism due to their high interdependency. PTMs can potentially affects any TFs' functional behaviour and many changes in behaviours, such as dimerization and protein-protein interaction, are prerequisite for bestowing determined functions.

TF mode of regulation

Regulation of TF synthesis is one of the most direct ways to control the expression of a defined set of genes in a restricted time and space. Tissue specific factors like the MyoD were shown to be essential and sufficient for embryonic-muscle fibroblast differentiation into myoblasts (R. L. Davis et al. 1987). Like other

myogenic regulatory factors, MyoD is expressed and synthesised exclusively in skeletal muscle. A further example of synthesis-driven regulation of TFs' activity comes from the homeobox transcription factor family (Hox genes). This family of transcription factors are used across several bilateral animals controlling developmentally regulated genes (Pearson et al. 2005). Mutation in Hox genes results in morphological defects restricted to certain segmental zones across the anterior-posterior axis. Hox genes are grouped in large collinear genetic clusters (Pearson et al. 2005) and their spatial organisation within the cluster reflects the order in which they are expressed (McGinnis & Krumlauf 1992). It is in fact the expression gradient of these genes that controls the precise expression of downstream "realisator genes", which harbour at their promoters high affinity or low affinity consensus for the Hox protein (Small et al. 1992).

TF concentration can also be controlled through degradation, as is the case with AP-1, HIF-1 and p53. The best example comes from the p53 transcription factor. p53 is essential in the control of cell cycle and apoptosis (Lane 1992). In normal cells p53 is kept ubuquitinated by the Mdm2 ubiquitin ligase (Haupt et al. 1997) and its protein levels are kept low through proteosomal degradation (Maki et al. 1996). Challenging cells with different stresses such as DNA damage, hypoxia and heat shock, blocks MDM2 function, thereby increasing p53 levels.

PTMs are a further key mode of regulation directly affecting TFs' biochemical properties. PTMs are one of the most studied mechanisms in TF regulation as they directly or indirectly control most if not all TF functions (Filtz et al. 2014). Surface-initiated signalling pathways induce an array of PTMs that decorate TFs. Signalling pathway crosstalk provides an integrated response to cues (Bardwell et al. 2007). Ligand - receptor - transducer is the typical path through which signal is delivered and amplified. Most studied PTMs affecting TFs include, among others, phosphorylation, sumoylation, ubiquitilation, acetylation and methylation (see Table 1.2).

Among the PTMs listed above phosphorylation is the most studied and consists of the addition of a phosphate group on selected amino acid residues by active kinases. Phosphorylation and de-phosphorylation can affect TFs' activity by altering their allosteric conformation and attractive/repulsive forces (Sprang et al. 1991). This change could affect several TFs' function. Protein stability of p53 and ATF-2 is increased upon phosphorylation (Fuchs et al. 2000; Appella & C. W.

Anderson 2000) while on the other hand TFs such as E2F-1 and MyoD are susceptible to degradation once they are phosphorylated (Song et al. 1998; Vandel & Kouzarides 1999). Negatively charged TFs, as mentioned in the description of the transactivation domain, constitute an effective interaction surface for several target proteins (see the VP16 TADs). The ability to induce an accumulation of negatively charged phosphate could constitute an effective mechanism to induce TADs activity.

Dimerization

Protein-protein interactions are employed by TFs in order to produce an enormous functional diversity (Klemm et al. 1998). TF dimerization provides an additional level where activator functions could be regulated. Dimerization relies on conserved motifs embedded into the TFs allowing them to form homotypic or heterotypic multimers (depending on whether proteins of the same or different family are interacting with each other). It is important to highlight that most if not all TFs are found as dimers or oligomers.

Diverse heterodimers are responsible for controlling specific sets of genes combining a limited number of TFs (Amoutzias et al. 2008). Considering all the possible pairs within each TF family, it is conceivable that dimerization could account extensively for flexibility and complexity in gene regulation but the functional regulation still has to be addressed (van Nimwegen 2003; Kummerfeld & Teichmann 2006). TFs complexes obtained through dimerization potentially hold different DNA binding specificity than their constituents. The AP-1 transcription factor family is a clear example. Composed of Jun, Fos and ATF, which can homo- or heterodimerize, the AP-1 family is able to generate different regulatory functions (reviewed in Karin et al. 1997 and Shaulian & Karin 2001). For example Jun-Jun or Jun-Fos dimers bind DNA elements containing TGACTCA sequence, whereas Jun-ATF or ATF-ATF associate with the DNA sequences TGACGTCA. However, the significance of the variation in formation of AP1 complexes for target gene selection has not been investigated.

Dimerization can also be induced upon signal transduction, in many cases controlling DNA binding affinity. In this circumstance dimerization plays a crucial role in acutely changing TFs' functional behaviours. For example, the JAK-

dependent phosphorylation of a single tyrosine residue at the C-terminus of STAT molecules induces their dimerization and accumulation into the nucleus (T. Meyer & Vinkemeier 2004). Similarly, Smad proteins undergo oligomerization upon TGF-β stimulation and concomitant nuclear retention (see within the following section).

Subcellular localisation

Cytoplasmic retention of TFs, away from their nuclear "playground", is an additional way to regulate them. This mechanism is especially important for signal-dependent transcription where effector proteins need to be maintained in an inactive state until a signal promotes their activation and relocation into the nucleus (Ziegler & S. Ghosh 2005).

This mechanism could be achieved by masking nuclear-localisation signal (NLS) within TFs. This unidirectional model was proposed for several TFs and co-regulators including, among others, NF-κB (whose NLS could be masked by IKBα), the steroid receptor (SR) and components of the Wnt signalling pathway (Htun et al. 1996; Moon & Kimelman 1998). Most if not all TFs controlled *via* subcellular localisation are found to be constantly shuttling between the cytoplasm and the nucleus (Ziegler & S. Ghosh 2005). Nuclear accumulation can therefore reflect a sophisticated regulation of import/export through differential binding interactions in the nucleus and in the cytoplasm. TF shuttling may ensure that inducible activators are inactive in the absence of signals and readily induced in the presence of signal. This mechanism will be further discussed in section 1.7 as a key mechanism controlling MRTF co-factor localisation/activity in the SRF transcription network.

One crucial finding has been the shuttling mechanism in NF-KB- IKBα localisation (Tam et al. 2000; Malek et al. 2001). Inhibition of the exportin CRM1 using Leptomycin B (LMB) induces NF-KB- IKBα nuclear accumulation, providing compelling evidence for an active import-export activity under resting conditions. This mechanism was described in more detail in several research papers, suggesting an imbalance between the contribution of import and export signals on NF-κB and IκBα (Tam et al. 2000).

Smad proteins represent another example of localisation-controlled transcriptional regulators. Smad transcription factors respond to transforming growth factor–β (TGF-β) ligands, including TGF-β and BMP (reviewed in Yigong

Shi & Massagué 2003). Smad proteins shuttle between the cytoplasm and the nucleus with discrete periods using NLS signals for import and CRM1-dependent or independent processes for their export (Pierreux et al. 2000; Xu et al. 2002). Following stimulation, phosphorylation-induced oligomerization of the Smad proteins promotes their retention in the nucleus allowing them to recruit co-activators and induce transcription (Inman & Hill 2002). Changes in shuttling behaviours after nuclear accumulation suggest a possible active retention of Smad proteins in the cytoplasm (Dong et al. 2000; Inman & Hill 2002).

TF shuttling is also a key mechanism in biological oscillations. Several biological processes could be described as rhythmic phenomena such as cell cycle progression, circadian clock and the activity of neuronal and cardiac cells (Belchetz et al. 1978; Shimojo et al. 2008; Gerber et al. 2013). It has recently been described in soil-living amoeba that nucleocytoplasmatic shuttling of TFs is essential to decode the number, rather than the level, of external stimuli ensuring developmental synchronisation of cells (H. Cai et al. 2014).

Protein-protein interaction

As described throughout the introduction, TFs can be regulated through protein-protein interactions. Co-regulator recruitment or multi-TF complex formation determines the function of most TFs allowing them to recruit, interact and regulate components of the transcriptional machinery. Beside direct recruitment of co-regulators, protein-protein interaction could also directly influence TFs' functional behaviours including nuclear localisation, DNA binding and avoid or favour other protein-protein interactions required for transcriptional activation. An example that will be described in more detail later is the regulation of MRTF co-factor nuclear localisation and activity (Miralles et al. 2003). In particular MRTF is able to interact with several actin molecules using specialized motifs at its N-terminus and these interactions have profound effects on MRTF behaviour and SRF-dependent gene activation (see section 1.7). Furthermore, as will be described in the results, MRTF-DNA association could also be controlled by its interaction with monomeric actin within the nucleus.

Recent works based on different organisms have tried to systematically dissect this intricate network of interactions. Recently Ravasi *et al.* generated an

atlas of combinatorial TFs, aimed at describing the way different combinations of TFs account for different modes of regulation (Ravasi et al. 2010). These studies highlight the importance of protein-protein interaction in TF regulation as key in determining cell fate. An interesting observation emerging from these TF networks is that interaction maps are evolutionarily conserved and can be traced between different organisms and species (Ravasi et al. 2010).

1.3 The transcriptional machinery

Pioneering work in 1959 led Samuel Weiss and Leonard Gladstone to discover the RNA polymerase as the primary enzyme responsible for DNA-dependent RNA synthesis in rat liver (Weiss & Gladstone 1959). At that time, the isolation from rat liver nuclei of this enzyme was not trivial. Since bacterial extracts were easier to make and handle, in 1965 Jacob, Monod and Lwoff succeed in the isolation of the first prokaryotic RNA polymerase. This research was awarded with the Nobel prize (Stent 1965).

Isolation from HeLa cell extracts (Roeder & Rutter 1970), rat liver (R. C. Conaway & J. W. Conaway 1990) and yeast (Lue & Kornberg 1987) showed a different scenario for eukaryotic cells. In contrast to bacteria, eukaryotic cells contain three different forms of RNA polymerase (I-III) (Roeder & Rutter 1970; Kedinger et al. 1970). As opposed to prokaryotes, eukaryote RNA polymerases are “blind” enzymes unable to initiate transcription given a dsDNA (Roeder & Rutter 1969; Roeder & Rutter 1970). Enzymatic activities were observed for less purified polymerase samples (Weil et al. 1979) suggesting that for eukaryotic polymerases additional factors were required. Several studies over the years showed that additional factors required for transcription *in vitro* are required for sequence specific initiation on dsDNA (Matsui et al. 1980). These are the so-called General Transcription Factors (GTFs) as they are generally required for the expression of all genes (Chambon 1975). The five major GTFs (TFIIB, -D, -E, -F, and -H) that assist the RNA polymerase lead to the recognition of the transcriptional start site (TSS) of a gene, melting of the double stranded DNA, and copying one strand (coding or leading) into RNA using ribonucleoside triphosphate and lastly re-winding the two strands of the DNA at its back while the polymerase translocates forward (see later in this chapter). This set of six proteins was still not sufficient to

explain how eukaryotic TFs were activating transcription. Studies aiming to recapitulate *in vitro* the mechanism of activated transcription lead to the identification of intermediary proteins, part of the Mediator complex (Y. J. Kim et al. 1994). The Mediator is highly sophisticated and complex machinery recruited by TFs and required to specifically potentiate initiation (Thompson & Young 1995; Takagi & Kornberg 2006).

The next sections summarise the key works that over the past decades shed more light on the complexity of the transcription process.

1.3.1 The DNA template

The correct positioning of the transcriptional machinery is crucial for accurate transcription of all genes. Each gene displays a so-called core promoter sequence composed of DNA elements. From textbooks we learnt that promoters are regions directing accurate transcriptional initiation by RNA polymerase II. It is possible to distinguish two kinds of promoters: the so-called “focused promoters”, showing a unique TSS and “broad promoters” where transcription initiation can occur at multiple sites (Juven-Gershon & Kadonaga 2010). As shown by Carninci *et al.*, focused promoters account for a minority in vertebrates (Carninci et al. 2006). Most promoters could be defined as ‘broad’ promoters showing more than one TSS linked to CpG islands around constitutively expressed genes. Several types of cis-elements contribute to core promoters’ activity (Juven-Gershon & Kadonaga 2010) and figure 1.1 shoes the consensus that has been catalogued through several studies based on single genes or hybridisation-based methods.

Inr: The Inr element spans the TSS and is probably the most common feature in core promoters (FitzGerald et al. 2006; Gershenzon & Ioshikhes 2005). Inr consensus varies according to organism and analysis, at mammalian promoters it has the sequence YYANWYY (IUPAC annotation) (Juven-Gershon & Kadonaga 2010). Conventionally the A nucleotide within the Inr motif is assigned to the +1 position as it generally represents the first nucleotide of the RNA.

TATA box and BRE: The TATA box was the first eukaryotic promoter element to be identified (Figure 1.1) (Lifton et al. 1978). As will be discussed later

the TATA element is recognised by the TBP subunit of the TFIID complex. This interaction is conserved from Archea to human (Reeve 2003). A perfect TATA consensus only exists in 10%-15% of core promoters (Carninci et al. 2006). A further conserved element is the TFIIB regulatory element or BRE (Lagrange et al. 1998). BREs are found both upstream or downstream of the Inr referred to as BREu for upstream and BREd for downstream (Figure 1.1).

DPE and MTE: The DPE, which is also contacted by the TFIID complex, is located downstream of the Inr (+28 to +33 from the A+1 of the Inr) and is conserved in Drosophila and human (Burke & Kadonaga 1997). Inr and DPE act cooperatively allowing TFIID to bind on certain promoters in a TATA-independent manner (Kutach & Kadonaga 2000). Upstream of the DPE an MTE element was found in Drosophila and human, and it can also cooperate with the Inr to locate the TFIID complex in a TATA-independent manner (Ohler et al. 2002; Lim et al. 2004). However MTE and DPE as well as MTE and TATA could act synergistically.

Although most promoters fall into the 'broad' category the distribution of cis-elements has shown to be conserved for most yeast and human genes (Rhee & Pugh 2012). Core promoter elements have very little or no tolerance for variable spacing, reflecting the structural constraints of the initiation complex (Kostrewa et al. 2009). Several maps of TSSs using cap analysis of gene expression (CAGE) still confirm the division in 'sharp' and 'broad' promoters as observed by Carninci in 2006 (Carninci et al. 2006; Juven-Gershon et al. 2006; FANTOM Consortium 2014). Recent studies in zebrafish tried to address how TSSs are selected and used throughout early embryonic development (Haberle et al. 2014). Analysis of maternal TSSs suggested that broad promoters could be constituted by multiple individual sharp TSSs. In order to acquire a definitive picture of promoter usage and TSS selection, more work needs to be done combing TSS maps and genome-wide transcriptional machinery footprints.

Besides the elements constituting core promoters, further elements have been identified as being essential to modulate the transcriptional output of a given gene. Early studies of the HSV promoter led to the identification of DNA elements in addition to the conserved TATA sequence required for maximal promoter activity (McKnight et al. 1981). It was following the works of the Schaffner lab that the term

enhancer was first coined (Banerji et al. 1981). The key characteristic that uniquely describes enhancer elements is that they can act irrespectively of the distance or orientation to the target promoter (Shlyueva et al. 2014). In addition to enhancers, silencers, insulators and tethering elements also play a crucial role in gene activation and, like the enhancers, are all distal elements of the TSS (Petrykowska et al. 2008; Gaszner & Felsenfeld 2006; Calhoun et al. 2002). Among these elements, enhancers are crucial for transcriptional activation. Enhancer activity relies on the combined action of different TFs and similarly gene activation requires interplay between several enhancers (Spitz & Furlong 2012). Enhancers' activity and their associated TFs are cell-type-specific (Bulger & Groudine 2011). These DNA elements are often devoid of nucleosome and correlate with a set of histone modifications such as histone H3 lysine 4 monomethylation (H3K4me1) and H3K27 acetylation (H3K27ac) (see in Chapter 1.4). The number of enhancers in the mammal genome has been estimated between 400,000 and 1.4 milion (Dunham et al. 2012; Thurman et al. 2012) but only a few thousand are used in a cell-type specific manner (Dunham et al. 2012; Yip et al. 2012; Y. Shen et al. 2012). Being able to identify all the enhancers contributing to the activity of a given gene is crucial in order to understand how activated transcription works.

1.3.2 Early transcriptional events

Pol II is the key enzyme of the transcriptional machinery that catalyses the synthesis of pre-mRNA, snRNA and miRNA (Cramer et al. 2008). Its 12 subunits (RPB1-RPB12) are highly conserved in all eukaryotes (Cramer et al. 2008). The main core of the RNA polymerase II, RPB1, RPB2, RPB3 and RPB11, represents a conserved structure in Pol I, Pol III, bacterial RNA polymerase and archeal polymerase (for a review see Cramer et al. 2008).

In higher eukaryotes, before heading to productive transcription, the transcriptional machinery has to overcome five major transitions (Figure 1.2) (Michel & Cramer 2013). Initially the assembly of the Pre-Initiation Complex (PIC), comprising Pol II and six GTFs (TFIID, TFIIA, TFIIB, TFIIE, TFIIF and TFIIH), occurs at dsDNA promoters. Before transcription can occur two further transitions are required (Figure 1.2). First the PIC isomerizes into an open complex where the promoter is partially melted. DNA melting consists of a so-called 'bubble', allowing

access of the initiating nucleotide to the template through the catalytic site. The next stage sees extension of the 'bubble' allowing RNA synthesis in a discontinuous fashion entailing for abortive transcript. Pol II escape from core promoters is achieved when a stable RNA-DNA hybrid is formed.

In vitro studies carried-out by several labs lead to a stepwise assembly pathway of the PIC obtained using purified GTFs and pure Pol II (Figure 1.3).

Everything starts with TFIID. TFIID is a key player in early transcription. Composed of the TATA-binding protein (TBP) and TBP associated factors (TAFs), TFIID is responsible for promoter recognition and Pol II positioning in an activator dependent way (Mencía et al. 2002). How TBP and TFIID recognise core promoters and how this complex recruits the Pol II machinery have been a matter of study for many years. The recognition of the TATA motif by TBP *in vivo* is not the only way TFIID is recruited. TBP was found also at 'TATA-less' promoters where evolutionarily conserved TFIID subunits such as TAF1 (Basehoar et al. 2004) and TAF4-12 (Gazit et al. 2009) contribute to promoter recognition. In addition TFIID can be also recruited to defined histone modifications. As will be presented in chapter 1.4, H3K4me3 is a hallmark of active promoters and TAF3, with its PHD finger domain can recognise this mark (van Ingen et al. 2008).

TFIID action is diverse as each TAF can have promoter-specific function (Ohtsuki et al. 2010). For example knockdown of TAF10, a subunit known to disassemble the TFIID complex, blocks embryonic liver development but has little effect on gene expression in adult adipocytes (Tatarakis et al. 2008).

TFIID-DNA interaction can be affected *via* several negative effectors such as Mot1 and NC2. *In vitro* and cell based studies have shown that TFIIA is essential to stabilise TFIID-DNA interaction and overcome these inhibitory effects, enhancing PIC assembly (Thomas & Chiang 2006).

The next GTF to enter the PIC is TFIIB. TFIIB is essential for further stabilisation of the TFIID-DNA complex and for Pol II/TFIIF recruitment (Cramer et al. 2008). TFIIB forces Pol II to assume an ordered conformation and concomitant binding of a second Mg atom in the Pol II active site (Sainsbury et al. 2013). Furthermore TFIIB, interacting with both TBP-DNA and the Pol II, controls the DNA path in the Pol II cleft (H.-T. Chen & Hahn 2004).

Pol II enters the PIC along with TFIIF, which may play an allosteric role in opening the Pol II clamp and ordering the DNA over the cleft (Y. He et al. 2013).

The positioning of the DNA is accompanied by its bending by ~18 degrees, a conformation that will be maintained in the complete PIC. These changes in conformation may facilitate TSS selection downstream of the BREd element (Y. He et al. 2013).

Finally TFIIE gets recruited to the PIC. Its major role, together with TFIIH (see later in this section), is to induce the open complex transition prior to promoter escape (see Figure 1.2) (Fishburn & Hahn 2012).

Composed of ten subunits, of which three have enzymatic activity, TFIIH is one of the most complex GTFs. The ATP-dependent enzymatic subunits are CDK7, XPD and XPB (Egly & Coin 2011). The helicase-like XPB promotes ATP-dependent unwinding of the DNA, characterising the first step in the transition from the PIC to the Open Complex (OC) state (Tirode et al. 1999). The XPB enzyme works as an ATP-dependent DNA translocase moving 15 bp of dsDNA towards the Pol II active site, leading to DNA unwinding (T. K. Kim et al. 2000). Consistent with this, recent maps of TFIIE and TFIIH binding locate them downstream of the start site (Grünberg et al. 2012; Y. He et al. 2013) Holding of TFIID, TFIIB and Pol II onto the upstream part of the promoter while TFIIH and TFIIE pump DNA into the active site, provokes the melting of the DNA strand in the cleft (Y. He et al. 2013). Recently it has been shown that this step is regulated in eukaryotes (Kouzine et al. 2013). Pol II is pre-loaded but DNA melting occurs only during activation, correlating with increases in TFIIH recruitment.

The OC is accompanied by an initial synthesis of RNA involving ‘scrunching’ of the DNA as the DNA cannot flow freely through the PIC while downstream DNA is pulled into the active cleft (Kapanidis et al. 2006; Revyakin et al. 2006). This step involves cycle of abortive transcription where the polymerase maintains contacts with the promoter and the GTFs (Kireeva et al. 2000). This step in part correlates with TSSs selection where the B-reader domain of TFIIB, interacting with the upstream DNA, plays a crucial role (Y. He et al. 2013).

The abortive phase is resolved once the Pol II synthesises a longer RNA chain leading to disassociation of TFIIB from Pol II (Y. He et al. 2013; Pal et al. 2005). This final step marks the end of the early transcription phase. Dramatic changes characterised by bubble collapse transition, promoter clearance and escape, will lead the Pol II in productive transcription.

1.3.3 Transcription activation: the Mediator complex

Since the discovery of the major components of the PIC it was clear that under suitable conditions this complex could suffice for basal transcription *in vitro*. Early models of activator function proposed that activation domains directly facilitated Pol II recruitment in order to stimulate transcription (Ptashne & Gann 1997). Failures in recapitulating activated transcription *in vitro* suggested that possible intermediary proteins were necessary. A breakthrough came in 1994 with the purification of the Mediator complex and the reconstitution of activated transcription *in vitro* (Y. J. Kim et al. 1994). This discovery demonstrated that recruitment of co-activator proteins by specific TFs is a major step in activated transcription.

Of all the co-activators identified over the past years, the 30-subunit Mediator complex is potentially the most crucial, allegedly having a widespread role across the transcriptome (Takagi & Kornberg 2006).

The Mediator complex is composed of 30 or more subunits that can be divided into four main modules called 'head', 'middle', 'tail' and 'kinase' (Figure 1.4) (G. Cai et al. 2009). The head, middle and tail modules form a stable 'core' while the kinase module associates reversibly with the Mediator complex (Malik & Roeder 2010). This transient association of the kinase module was suggested to correlate with inactive to active state transition (Malik & Roeder 2010). The overall structure of the Mediator complex is heterogeneous. Incorporation of paralogues subunits and tissue-specific subunits makes this complex highly diverse (Sato et al. 2004). Although the modules and overall structure are conserved across eukaryotes, several subunits diverge significantly in mammals when compared to yeast (Bourbon 2008). Some conserved residues, located in structurally disordered domains, are potential interfaces acquiring structural features upon interaction with other proteins (Tsai et al. 2014).

As mentioned earlier the Mediator complex could be considered a major end point for signalling pathways (Figure 1.4). Activated nuclear receptors, the MAPK-regulated factor Elk1 and the regulator of lipid metabolism SREBP1α are a few examples of TFs interacting with different subunits of the Mediator complex (P. Jiang et al. 2010; G. Wang et al. 2005; F. Yang et al. 2006). Interactions between TFs and Mediator complex show extensive architectural rearrangement of the latter

(Bernecky & Taatjes 2012). These rearrangements were shown to be required for transcriptional activation as much as TFs recruitment and PIC formation (Meyer et al. 2010).

Mediator's function as a bridging complex becomes clear with works aimed at understanding the dynamic interaction between promoters and distal elements. Studies focused on MED1 and the nuclear receptor (NR) interaction have shown that knockdown of MED1 negatively regulates NR-dependent genes followed by a loss of looping between enhancers and promoters (Park et al. 2005). The Mediator complex was found to work cooperatively with cohesin to establish enhancer-promoter loops (Kagey et al. 2010). Moreover the distribution of these loops, and the occupancy of Mediator and cohesin, changes with cell differentiation (Kagey et al. 2010).

Each module of the Mediator domain exerts different functions. Most Mediator-TFs interactions were mapped within or in close proximity to the tail module (Malik & Roeder 2010) while most of the functional activities occur at the head module (Figure 1.4).

As mentioned earlier the Mediator complex bridges TFs with the PIC. Indeed Mediator is able to interact with most of the Pol II subunits and in particular with the unstructured corboxy-terminal domain (CTD) (Robinson et al. 2012; Soutourina et al. 2011). Upon interaction the Mediator head module undergoes extensive structural shifts (Naar et al. 2002). Together with Mediator's structural changes, Pol II undergoes conformational variations such as a possible movement of the clamp, as suggested from recent EM studies (G. Cai et al. 2009). Furthermore the Mediator complex is allegedly able to interact with all the GTFs composing the PIC. Works done by the Carey, Roeder and Conaway labs have shown that the Mediator can interact and specifically recruit TFIID to promoters (K. M. Johnson et al. 2002; Guermah et al. 1998; Takahashi et al. 2011). Additionally the Cramer and Werner labs demonstrated direct interaction between the Mediator subunit MED11 and TFIIH (Esnault et al. 2008; Seizl et al. 2011). This interaction enhances TFIIH kinase activity towards Pol II CTD (see later in this chapter).

Beside its clear role in PIC formation, Mediator also plays a crucial role in post-recruitment steps both at Pol II pause-release and transition to active elongation. A full description of these functions will be presented in specific sections throughout the introduction.

1.3.4 Polymerase pausing during early elongation

As described previously, PIC formation is a major step in controlling RNA transcription. A further regulatory block in the transition to productive transcription of metazoan genes is the promoter-proximal Pol II pausing (Adelman & Lis 2012) (Figure 1.3). Pol II pausing is an intermediate step between promoter escape and productive transcription, suggested to provide a possible input for transcriptional regulation. This transition occurs when Pol II is already stably engaged, detached from TFIIB and the chain of RNA is longer than 12 nts (Kwak & Lis 2013). Paused Pol II describes a stage in the transcription cycle different to Pol II arrest. During elongation Pol II is susceptible to arrest where Pol II is found in a backtracked position with the nascent RNA spanning the active site (Bengal et al. 1991). In this circumstance TFIIS recruitment is required to stimulate Pol II to cleave the protruded RNA, re-establishing a realigned RNA 3' end with the Pol II active site (Bengal et al. 1991). TFIIS activity on arrested Pol II is key in promoting efficient Pol II elongation (see Chapter 1.3.5).

Pol II pausing was first observed on chicken and human genes where Pol II levels at gene promoters were significantly higher than in the gene body (Gariglio et al. 1981; Gilmour & Lis 1986; Krumm et al. 1992). Nuclear run-on experiments showed that this Pol II was transcriptionally engaged (Giardina et al. 1992). Use of high salt or ionic detergent (sarkosyl) is essential for the paused Pol II to transcribe, indicating that removal of chromatin proteins is indispensable in these assays (Rougvie & Lis 1988). Paused Pol II was shown to be associated with the DRB (5,6- dichloro-1-β-D-ribofuranosylbenzimidazole) sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) (Wada et al. 1998; Yamaguchi et al. 1999). DRB is a small molecule known to inhibit most transcriptional CDKs with a higher affinity for CDK9, the kinase part of the p-TEFb complex (see later within this chapter). Several hypotheses have tried to model how DSIF and NELF are able to induce pausing but definitive evidence is still missing. Genome wide analyses have shown that both NELF and DSIF associate with Pol II at promoter proximal regions while only DSIF is also distributed along the gene body where Pol II elongates (Rahl et al. 2010). In higher eukaryotes there is some evidence that RNA binding factors could determine pausing. Both NELF and DSIF were shown to interact with short RNA emerging from the elongating Pol II (Fujinaga et al. 2004;

Narita et al. 2003; Missra & Gilmour 2010). Other hypotheses are based on the 'barrier' model. As described in a recent review by Kwak and Lis, the first nucleosome could provide an energy barrier that needs to be overcome *in vitro* (Kwak & Lis 2013). This model is supported by the observation that in Drosophila the first nucleosome interacts with Pol II (Mavrich et al. 2008).

Paused Pol II was shown to provide a lapse of time for different checkpoint mechanisms to occur (Figure 1.3). The first checkpoint is provided by the 5' Capping of the RNA. As will be discussed later, transcription is accompanied by other essential events that contribute to proper processing of the RNA (see chapter 1.3.7). In metazoan a unique enzyme is able to catalyse two major modifications of the RNA 5' end. Capping enzyme (CE) digests the 5' triphosphate of RNA and is able to add a guanine base (Bentley 2014). When the Pol II is ready to head into elongation CDK7 Kinase phosphorylates the Pol II CTD on Serines 5 (see chapter 1.3.6). This PTM recruits CE that, interacting with Pol II and DSIF, relives the action of NELF providing a platform for p-TEFb (Lenasi et al. 2011; Mandal et al. 2004). P-TEFb complex, a heterodimer of the kinase CDK9 and Cyclin T, is essential for productive transcription and can be recruited in several ways to the paused Pol II (D. H. Price 2000). In human cells p-TEFb is maintained in a repressive complex by the 7SK snRNP (small nuclear ribonucleoprotein) (Nguyen et al. 2001; Z. Yang et al. 2005). Upon activation p-TEFb is released and relocated to promoters. Recruitment of p-TEFb leads to phosphorylation of both NELF and DISF (Yamada et al. 2006). These modifications allow Pol II to detach from NELF and enter into elongation in association with DSIF. The CE-induced and the p-TEFb dependent mechanism of Pol II pause release might act either synergically or in a context specific manner. Further insights are required in order to gain a full picture of this mechanism.

TFs such as c-MYC, NF-κb and the human immunodeficiency virus (HIV) TAT trans-activator were shown to directly interact with p-TEFb (Eberhardy & Farnham 2002; Barboric et al. 2001; D. H. Price 2000). Co-regulators can also recruit p-TEFb. BRD4 is a co-activator containing a bromodomain able to directly interact with p-TEFb (Z. Yang et al. 2005; Jang et al. 2005). By means of its bromaodomain BRD4 is able to recognise acetylated histones that mostly occur at gene promoters. As proposed by the Oliviero lab, BRD4 can be recruited as a result of a histone modifications cross-talk leading to H4K16 acetylation (see

chapter 1.4) (Zippo et al. 2009; Zippo et al. 2007). Although it is not yet clear how this histone cross talk is initiated and maintained at gene promoter, BRD4-dependent CDK9 recruitment seems crucial for *Fosl1* gene activation.

Paused Pol II is suggested to bring two advantages for gene activation. It was shown in several systems that having paused Pol II at gene promoters allows prompt and synchronized responses to activating signals (Bentley & Groudine 1986; Kao et al. 1987; Rougvie & Lis 1990; Gilmour & Lis 1986). Furthermore pausing factors such as NELF are suggested to play a positive role in maintaining a permissive landscape at defined promoters (Muse et al. 2007; Gilchrist et al. 2008). Reduction of pausing factors such as NELF, negatively affects transcription of several genes in mouse embryonic stem cells (mEsc) (Rahl et al. 2010; Gilchrist et al. 2008).

1.3.5 Polymerase elongation and processivity

The action of p-TEFb is restricted at promoter proximal Pol II as it is not required once productive elongation has started (Y. Jiang et al. 2004; Egyházi et al. 1996). Several changes occur on the traveling polymerase starting with the loss of NELF and retention of DSIF after phosphorylation (Figure 1.3). One major observation is the change in traveling or elongation rate. Paused Pol II is moving less than 5 bases per minute, being essentially static (Peng et al. 1998) while elongating Pol II travels at ~3-4 kb/min, depending on the gene (Danko et al. 2013). The speed of the transcribing Pol II is not homogeneous along the template as several obstacles through its path can interrupt it. Promoter proximal pauses, termination, folding of nascent RNA, polyadenylation, splice site selection, the nucleosomes or even clashing with other polymerases are a few examples affecting the elongation rate. Several factors have been described as involved in facilitating Pol II processivity and elongation rate.

TFIIS and TFIIF were the first two factors discovered with a role in Pol II elongation. As described earlier, TFIIS activity is crucial to liberate Pol II from arrested sites along the template both *in vivo* and *in vitro*. Genome-wide studies conducted in Drosophila showed that TFIIS function is exerted at early elongation complexes (Nechaev et al. 2010). Beside its role in initiation, TFIIF is also involved in Pol II elongation. TFIIF increases the elongation rate by reducing Pol II dwell-

time (the time Pol II spends at each successive nucleotide) (D. H. Price et al. 1989; Flores et al. 1989). TFIIF does not remain associated with Pol II throughout elongation and its activity *in vitro* is concentration dependent, allowing a 20-fold increase in elongation rate.

A further complex involved in pause release and enhancement of Pol II catalytic rate is the super elongation complex (SEC) (Lin et al. 2010). This complex was initially identified through studies of aberrant gene activation in mixed lineage leukaemia (MLL) (Luo et al. 2012). SEC can be described as an assembly of different factors exerting diverse functions, including AFF1 or AFF4 (AF4/FMR2 family members), ENL (elevated-ninteen leukemia) or AF9 (ALL fused gene from chromosome 9), ELL1 or ELL2, and p-TEFb (Luo et al. 2012). Although SEC complex has a clear role in promoting elongation both *in vitro* and *in vivo* much work is needed to elucidate the biochemical mechanism entailing its activity.

As mentioned earlier chromatin is a major obstacle to the travelling Pol II. Even after release of the Pol II from the PIC the transcribing machine requires factors helping in removing or weakening the nucleosome barrier. The FACT (facilitates chromatin transcription) complex was identified from HeLa nuclear extract as a crucial component allowing *in vitro* transcription from chromatinised templates (Orphanides et al. 1998). Made of SPT16 and SSRP1 subunits, FACT is able to evict H2A/H2B dimer from the nucleosomes allowing Pol II to go through the histone hexamer (Belotserkovskaya et al. 2003). As Pol II transcribes, FACT is able to replace back the H2A/H2B dimer (Xin et al. 2009).

A further factor involved in promoting Pol II elongation rate is the histone chaperon SPT6 (Yoh et al. 2008). Due to its ability to interact with histone proteins it was proposed to facilitate elongation *via* nucleosome destabilisation (Bortvin & Winston 1996). Beside its possible role in modifying the chromatin structure, SPT6 also affects Pol II elongation rate *in vitro* on naked DNA (Endoh et al. 2004). Furthermore in mammals SPT6 was reported to interact with Ser2 phosphorylated Pol II (see chapter 1.3.6). This interaction is required to recruit chromatin remodelling complexes and mRNA export factors (M. Sun et al. 2010; Yoh et al. 2007).

The PAF1 (polymerase associated factor 1) complex is a further factor having a role in Pol II elongation. PAF1 complex was first identified in yeast as a Pol II-associated factor (X Shi et al. 1996). PAF1 complex is composed of Ctr9,

CDC73, Rft1, Leo1, Ski8 and Paf1 (Jaehning 2010). PAF1 co-localises with Pol II from the promoter to the 3' end of genes (M. Kim et al. 2004; Mayer et al. 2010). During the transcription cycle PAF1 mediates several events including co-transcriptional histone modifications (Krogan et al. 2003), Pol II CTD phosphorylation and recruitment of termination factors (Nordick et al. 2008; Sheldon et al. 2005). Elegant biochemical studies have shown that PAF and DSIF are able, non-redundantly, to cooperate in stimulating Pol II elongation both *in vivo* and *in vitro* (Y. Chen et al. 2009). Furthermore recent studies from the Roeder lab have shown that PAF1 promotes transcriptional elongation through direct interaction with Pol II as well as TFIIS (J. Kim et al. 2010). Furthermore PAF1 was co-purified with components of the SEC complex leading to the hypothesis that PAF1 mediates the recruitment of the SEC complex to Pol II on chromatin (N. He et al. 2011).

1.3.6 Pol II CTD as platform coordinating the transcription cycle

RNA Pol II is unique among the three Polymerases in eukaryotes as it is equipped with an appendix – the Carboxy-teminal domain (CTD) within the Pol II large subunit. Pol II CTD is made of several heptad repeats showing the consensus sequence YSPTSPS (Eick & Geyer 2013). Pol II CTD coordinates the entire cycle of transcription from initiation to termination with a crucial function in pre-mRNA maturation. Beyond the production of RNA, Pol II CTD is involved in the epigenetic control of a cell contributing to "read", "write" and "erase" epigenetic marks.

The main feature of the CTD is the heptad repeat. This consensus is conserved in yeast and mammals. In yeast the CTD is composed of 26-29 heptad repeats while in mammals it reaches 52 repeats, with a greater degeneration of the consensus towards the N-terminal part (Eick & Geyer 2013). Deletion of a few repeats has little effect while removal of half or more of the CTD, in both mammals and yeast, is detrimental for cell viability (Thompson et al. 1993; Bartolomei et al. 1988). Extensive analysis done by Stiller and Shuman dissected the functional unit of the CTD in S. cerevisea and S. pombe. This unit is composed of an entire heptad repeat plus the first four amino acids (YSPT) of the next repeat, giving a final unit of YSPTSPS-YSPT (P. Liu et al. 2008; Schwer et al. 2012). In addition the

positions of the three serine-proline motifs and the two tyrosine residues were found to be crucial for CTD functions (P. Liu et al. 2008).

The Pol II CTD can be defined as one of the longest polypeptide chains without any charged residue. Serines, threonines and tyrosines with their hydroxyl-groups constitute a hydrophilic surface (Eick & Geyer 2013). The five amino acids composing the heptad repeat could all be modified by phosphorylation. By including dynamic glycosylation of some residues, proline isomerisation and the several PTMs there are 432 possible configurations for a heptad repeat and 10,368 for the functional unit (Eick & Geyer 2013). In the context of the full length CTD *in vivo* it is not clear how many combinations are possible. Many CTD modifications could affect each other accounting for a more convoluted system. Alongside the vast number of possible configurations, the CTD is a highly flexible and mostly unstructured domain with a length 5 to 6 times greater than the diameter of the RNA Pol II (Figure 1.5) (Cramer 2001). Electron micrographs of the Pol II in an un-modified state revealed weak density attributed to the CTD with a dimension of about 100 Å (Meredith et al. 1996). Furthermore, as shown in one of the first structures of the Pol II, the Pol II crystal has only a limited space close to the linker (Cramer 2001). This observation lead to the hypothesis that hypo-phosphorylated CTD is in a compact conformation. Extensive phosphorylation is a major cause of decompaction of the CTD due to charge repulsion (J. Zhang & Corden 1991).

Several kinases contribute to the modification of the Pol II CTD. Different complexes are responsible to deliver these kinases in a specific moment of the transcription cycle. CTD PTMs generate defined docking units characterised by a unique phospho-signature recognised by ‘readers’ of the Pol II CTD (Eick & Geyer 2013). Table 1.3 presents a list of CTD interacting proteins divided in to ‘writers’ such as kinases, ‘readers’ and ‘erasers’ of the CTD marks. As described earlier complexes such as TFIIH and the Mediator are responsible of delivering certain kinases in defined moments of the transcription cycle. Complex-delivered kinase is not the only mechanism as the CTD itself could provide docking sites recognised by kinases. This mechanism allows an ordered pattern of phosphorylation detected across the transcription unit. Similar mechanisms could occur for phosphatases that also contribute to the shaping of the CTD code.

In order to allow initiation Pol II has to be hypo-phosphorylated (R. C. Conaway et al. 1992). Phosphorylated CTD is unnecessary for PIC formation but is

indispensible to activate transcription (R. C. Conaway et al. 1992). Indeed phosphorylation of the Pol II CTD is essential to overcome negative effects caused by the interaction with the Mediator complex that avoids Pol II entering into elongation (Wong et al. 2014). The structure of Mediator and Pol II CTD has recently been described, elucidating the current model where the CTD acts at initiation to favour interaction with the PIC and Madiator but subsequently becomes detrimental (Wong et al. 2014; Robinson et al. 2012).

One of the first kinase to enter into the transcription cycle is CDK7 with TFIIH. CDK7 was shown both *in vitro* and *in vivo* to phosphorylate the CTD heptad repeats at Ser5 and Ser7 (Shiekhattar et al. 1995).

A second kinase, part of the Mediator complex, is CDK8. It was reported that CDK8 could phosphorylate Ser2 and Ser5 residues with negative effects in re-initiation *in vivo* (for a recent review see Nemet et al. 2014). Works from the Reinberg lab have shown that CDK8 can negatively affect transcription. Recombinant CDK8 is able to phosphorylate the Cyclin H subunit of TFIIH such that TFIIH activity is impaired (Akoulitchev et al. 2000). Its role in phosphorylating the CTD is directly dependent on its association with the Mediator complex, as recombinant CDK8 shows no activity towards Pol II CTD *in vitro* (Nemet et al. 2014). Recent works conducted by the Espinoza lab showed that CDK8 is also required for activation of several genes by targeting several TFs (Donner et al. 2010). Thus CDK8 may act in a context specific manner. Further investigations regarding the CDK8-dependent control of gene activity and Pol II phosphorylation are needed.

As mentioned earlier CDK9 is another kinase that is recruited at the promoters of active genes. CDK9 was thought to be involved in Serine 2 phosphorylation of Pol II as DRB treatment was clearly impairing Pol II elongation (Rahl et al. 2010). Recent studies have shown that CDK9 preferentially phosphorylates Ser5 and partially Ser7 *in vitro* (Baumli et al. 2008; Czudnochowski et al. 2012). In these studies human p-TEFb activity appears to be enhanced on a Ser7 pre-phosphorylated peptide. This result was also confirmed in yeast where Ser7-P seems to prime CDK9 activity (St Amour et al. 2012). *In vivo* live imaging further elucidates the co-localisation between CDK9 and Ser5 but not Ser2 (Ghamari et al. 2013). Although mammal CDK9 does not seem to contribute to Ser2 phosphorylation the co-regulator BRD4, found in complex with CDK9, seems

to phosphorylate Ser2 through a noncanonical N-teminal kinase domain (Devaiah et al. 2012). These results require confirmation and a more accurate biochemical analysis.

The identification in higher eukaryotes of kinases involved in elongation has been a matter of debate. The recent discovery of CDK12/CDK13 seems to have solved a longstanding enigma (Bartkowiak et al. 2010). In S. pombe and S. cerevisea two kinases are involved in phosphorylation of Pol II at initiation and elongation. Bur1 in S. cerevisiae and CDK9 in S. pombe are involved in phosphorylating Pol II and DSIF (NELF is a metazoan specific factor) while Ctk1 in S. cerevisea and Lsk1 in S. pombe are contributing to Ser2 phosphorylation and therefore efficient elongation (Johnsen 2012). In metazoan CDK12 could be the main kinase contributing for Ser2 phosphorylation of Pol II CTD towards the end of the gene body. As described by Bosken and collaborators, CDK12 activity is enhanced when the Pol II CTD is phosphorylated at Ser7 (Bösken et al. 2014).

Alongside the phosphorylation of Serine residues in the functional unit of the CTD, Tyrosine and Threonine could also be phosphorylated. Thr-4P seems to be recognised mainly by 3' end processing factors on chicken histone genes (Hsin et al. 2011). In mammals Polo-like kinase 3 (Plk3) seems to be a specific Thr-4 kinase (Hintermair et al. 2012) Tyr1 seems to be phosphorylated by Abl1 and Abl2 in human cells (Baskaran et al. 1997). This modification in yeast plays a crucial role in termination (see chapter 1.3.8).

The defined pattern of phosphorylation of the Pol II CTD is also determined by phosphatases. Phosphatases are not only crucial in determining the CTD phosphorylation status but also to prepare Pol II for re-initiation. Two major phosphatases are involved in this process, named Fcp1 and Ssu72. Fcp1, conserved both in human and yeast, exerts its action during elongation and is required for Pol II re-initiation (Chesnut et al. 1992; Archambault et al. 1997; H. Cho et al. 1999). Recent studies have shown that Fcp1 is mainly found at the 3' end of active genes and its inactivation leads to Ser2-P accumulation (D. W. Zhang et al. 2012a; Bataille et al. 2012). Ssu72 is a further conserved phosphatase that acts in de-phosphorylating both Ser5-P and Ser-7P (D. W. Zhang et al. 2012a; Bataille et al. 2012). Ssu72 was found as a component of the yeast cleavage and polyadenylation factor (CPF) and its activity is enhanced by interaction with components of the CPF (Dichtl et al. 2002; X. He et al. 2003; Nedea et al. 2003;

Ghazy et al. 2009). In HeLa cells Ssu72 phosphatase activity was found to be essential in 3' end processing suggesting that Ser5-P dephosphorylation is essential at the end of the gene (Xiang et al. 2010).

These and other phosphatases, such as the small CTD phosphatases (SCPs) in human, contribute to shape the Pol II CTD. Several genome-wide studies both in mammals and yeast contributed to describing the distribution of the CTD PTMs for an average gene (Figure 1.6). As described in Figure 1.6, signal for both Ser-5p and Ser7-P increases right at the TSS of active genes. While the signal for Ser7-P remains high throughout the transcription unit, Ser5-P signal slowly decreases towards the poly-A (pA) site. In contrast signals for Tyr1-P, Ser2-P and Thr4-P are low at the TSS but increase downstream in the gene body. At the crossing of the pA site the Pol II CTD presents a defined pattern that was shown recently to be crucial in yeast for certain termination events (Mayer et al. 2012). While Ser2-P reaches its maximum as it cross the pA site, Tyr1-P and Thr4-P drastically decrease as a result of phosphatase action.

1.3.7 Co-transcriptional RNA processing

Beside the regulation of the transcription process itself, CTD phosphorylation is allegedly fundamental in all RNA processing reactions. The CTD is the major platform required to couple transcription to RNA processing. Analysis of Pol II mutants in higher eukaryotes lead Bentley and colleagues to show for the first time that Pol II CTD is required for mRNA processing (McCracken et al. 1997). Pol II CTD is unnecessary for transcription per-se but is essential for 5' capping, splicing, and 3' end processing (McCracken et al. 1997).

Capping: The first step in mRNA maturation involves capping of the nascent RNA. As mentioned earlier this step correlates with TFIIH recruitment and disassociation of the Pol II from the pausing site (Adelman & Lis 2012) (Figure 1.3). The capping process consists of three main enzymatic activities, including triphosphatase, guanylyltransferase and methyltransferase (Cowling & Cowling 2009). Three enzymes in yeast are responsible for carrying out this reaction. On the other hand in mammals the triphosphatase and guanylyltransferase activities are performed by the bi-functional enzyme RNGTT while the addition of a methyl

group is performed by the methyltransferase RNMT (for review see Cowling & Cowling 2009)

Recently it was possible to observe both *in vitro* and *in vivo* that the capping reaction does not always reach completion and the step between RNGTT and RNMT activity can be controlled (Jiao et al. 2010). In particular, in yeast the Rai1 and Dxo1 enzymes are able to remove specifically 5'-end cap when unmethylated, and degrade the aberrant transcript providing a further quality-control mechanism in RNA processing (Jiao et al. 2010; Chang et al. 2012). This mechanism was recently shown to also apply in higher eukaryotes where Dom3Z is able to function on incompletely capped RNA (Jiao et al. 2013). Dom3Z as Dxo1 are both able to remove the cap structure, generating a 5' end monophosphate, and degrade the entire RNA with a 5'-to-3' exoribonuclease activity. As will be described later, de-capping can prime premature transcription termination due to the action of XRN2 5'-to-3' exoribonuclease (Brannan et al. 2012). Furthermore, as mentioned earlier co-transcriptional capping coincides with promoter proximal paused Pol II. Beside differences in architecture and structure across different taxa, the interaction between the guanylyltransferase enzyme and Pol II requires an intact CTD phosphorylated on Ser5 and Tyr1 (C. K. Ho & Shuman 1999; A. Ghosh et al. 2011).

Observations in S. pombe and mammals are suggesting a possible function of the capping in Pol II pause release. In S. pombe the completion of the cap is achieved with the recruitment of the methyltransferase Pcm1 that requires interaction with CDK9 in addition to the CTD (Viladevall et al. 2009; Pei et al. 2003). In higher eukaryotes it was also possible to determine a direct interaction between RNGTT and DSIF (Wen & Shatkin 1999). This interaction was reported to increase RNGTT activity. Furthermore data from the Reinberg lab brought these observations further as they reported that the capping enzyme is capable of reversing the negative effect of NELF on Pol II elongation, possibly by competing for Pol II interaction (Mandal et al. 2004). In addition cap-binding Protein Complex (CBC) was suggested to coordinate correct mRNA capping with p-TEFb activity (Lenasi et al. 2011). Furthermore, TFs such as Myc and the HIV TAT protein could induce RNGTT and RNMT recruitment, providing another possible layer of regulation (Chiu et al. 2002; Cowling & Cole 2010).

Splicing: In 1978 Walter Gilbert described eukaryotic genes as shredded into pieces, where the coding bits are intercalated by non-coding sequences called introns (Gilbert 1978). Most animals and plants present multiple introns per gene, whereas in fungi and unicellular eukaryotes introns seem to have drifted away (Rogozin et al. 2012). Despite these differences in intron density across eukaryote genomes, the highly sophisticated machinery responsible for intron removal, named spliceosome, is conserved. The spliceosome is a large ribonucleoprotein composed of five snRNPs (U1, U2, U4, U5 and U6) and hundreds of accessory proteins (Jurica & M. J. Moore 2003). Splicing and transcription can occur separately, but evidence has shown that co-transcriptional splicing is fundamental in transcription (Jeronimo et al. 2013). Transcription coupled splicing seems primarily controlled *via* the Pol II CTD. Several studies have shown how different components of the spliceosome are specifically recruited through the Pol II CTD (Eick & Geyer 2013; Hsin & Manley 2012). *In vitro* experiments demonstrated that hyper-phosphorylated Pol II is able to initiate several steps of spliceosome assembly (Hirose et al. 1999).

The assembly of the spliceosome is based on specific, consensus sequences on the growing RNA called splicing sites. Splicing sites are delimiting intron sequences both at the 5' and the 3' end. The first step in spliceosome assembly is the recognition of the 5' splice site by the U1 snRNP that is responsible for cleaving the 5' end of the intron (Jeronimo et al. 2013). The action of U1 snRNP allows the formation of a 'lariat' intermediate where the 5'-end of the cut intron attaches to a branch point through a pair of guanine and adenine forming a loop structure (for review see McManus & Graveley 2011). The next step consists of the recruitment of the U2AF snRNP to the polypirimidine tract of the 3' splice site so that the U2 snRNP can be recruited to the branch point in order to form the pre-spliceosome complex. Key subunits of U2AF were reported by the Manley and Bensaude groups to interact directly with the phosphorylated CTD (David & Manley 2011; Gu et al. 2013). This network of interaction involving snRNPs, the CTD and the RNA lead Manley and colleagues to propose that the CTD function consists of tethering the 5' splice site in proximity to the RNA exit channel of Pol II (David & Manley 2011). The newly transcribed 3' splice site, loaded with U2AF in a Ser2-P dependent manner, is going to be in close proximity to the tethered 5' splice site. This elegant model still needs experimental confirmation.

The transcriptional elongation rate and processivity influence the splicing efficacy due to a 'kinetic coupling' of the two processes (Eperon et al. 1988). Induction of a transcription pause can indeed favour the inclusion of exons (Roberts et al. 1998). The pause site in this circumstance allows more time to assemble the spliceosome leading to exon inclusion. As phosphorylation of the CTD has been reported to influence Pol II processivity, it is clear that CTD phosphorylation could indirectly influence splicing efficiency (Muñoz et al. 2009).

1.3.8 3' end processing and the transcription termination

Processing of the RNA 3' end is a key step in transcription. Due to the crucial role of the 3' end in transcription termination the two processes are going to be introduced together. The characterisation of the 3' untranslated region (UTR) started with the analysis of specific highly transcribed genes including the globin gene, the ovalbumin gene and the immunoglobin gene (for review see Proudfoot 2011). These pioneer studies lead to the identification of a long poly(A) tail synthesized by a poly(A) polymerase (Takagaki et al. 1988). Sequencing of six mRNA 3' ends led Proudfoot and colleagues to identify the common sequence AAUAAA 20-30 nt adjacent to the poly(A) tail (Proudfoot & Brownlee 1976). This sequence turned out to be indeed essential for 3' end poly(A) tailing. In addition, further elements appeared to contribute to a functional polyadenylation site (PAS) including a GU-rich sequence (downstream of AAUAAA also called DSE), and an U-rich tract (upstream of AAUAAA). Recent genomic studies conducted by Ozsolak *et al.* confirmed the generality of the PAS sequence and organisation (Ozsolak et al. 2010).

The 3' end processing consists in cleavage and polyadenylation of the nascent RNA in the correct position and at the right time during transcription. Manley and Keller performed extensive biochemical studies using HeLa cell nuclear extract in order to purify factors associated with the 3' UTR of a synthetic RNA substrate (Millevoi & Vagner 2010). Two protein complexes were purified: the cleavage and polyadenylation specific factor (CPSF) and the cleavage stimulatory factor (CstF). Both factors associate with the AAUAAA and the GU-rich DSEs promoting cleavage of the pre-mRNA. Furthermore the poly(A) polymerase was shown to be recruited to the 3' end of the cleaved product. The complexes involved

in what looks like a simple enzymatic reaction, cleavage and polyadenylation, are composed of several subcomplexes comprising 50 or more polypeptides. 3' end formation can be uncoupled to polymerase termination (C. L. Moore & Sharp 1985; Butler & Platt 1988). In addition cleavage and polyadenylation could be separated as two distinct reactions using ATP inhibitors or EDTA (Butler & Platt 1988; Zhao et al. 1999).

The idea that each step in 3' end formation is distinct and separable becomes clear with the study of 3'-end formation in replication-dependent histone mRNAs (Gick et al. 1986). Replication-dependent histone mRNAs do not have canonical PAS and these mRNA are formed through the recognition of the histone downstream element (HDE), a purine-rich sequence downstream of the conserved 3'-terminal hairpin, by the snRNA U7 (Schaufele et al. 1986; Schümperli 1988). The U7 snRNA, associated with several proteins constituting the U7 snRNP, is recruited to the 3' ends of histones mRNA by interacting with other factors recognising the 3' terminal hairpin (Marzluff et al. 2008). These, together with additional components of the cleavage/poly(A) complex, such as CPSF-73, CPSF-100 and Symplekin specifically recognise and cleave the 3' end mRNA.

Although the 3' end formation can occur both *in vitro* and *in vivo* independently of transcription termination, this latter mechanism requires 3' end processing (Whitelaw & Proudfoot 1986). Transcription termination is indeed coupled to 3' end processing as mutation of the PAS sequence allows Pol II to read through and bypass the normal termination site (Whitelaw & Proudfoot 1986; Connelly & Manley 1988). Recruitment of the cleavage/poly(A) complexes to the elongating Pol II requires an intact CTD (McCracken et al. 1997; Zhao et al. 1999). In particular phospho-CTD specifically interacts with both CPSF and CstF. A crucial example is the Pcf11 component of the CFII in mammals, and CF1A in yeast. Pcf11 is specifically recruited by Ser2-P and its distribution over the elongating Pol II is controlled by Tyr1-P that, at least in yeast, masks Ser2-P until the polyadenylation site appears (Mayer et al. 2012). Cleavage by the CFII complex in mammals, or the CF1A in yeast, exposes residual uncapped RNA attached to the elongating Pol II. Both in yeast and mammals this RNA is recognised and degraded by the 5'-3' exonucleolytic RNAase (Rat1 in yeast and XRN2 in mammals) (M. Kim et al. 2004; West et al. 2004). This exonuclease can be directly recruited to the elongating Pol II *via* p54nrb/PSF that can interact with both CTD and XRN2

(Kaneko et al. 2007). It is thought that the exonuclease is in kinetic competition with the elongating Pol II such that, when RNA degradation reaches the elongating Pol II, termination will be induced by conformational changes in the Pol II active site (Connelly & Manley 1988; Proudfoot 1989). This mechanism is called the Torpedo model and likely occurs in coordination with the cleavage/poly(A) factors ensuring correct and efficient termination of the transcribing Pol II. In order for the Torpedo mechanism to occur, the polymerase has to either slow down in a certain position after the PAS element or additional and more efficient cleavage has to occur proximal to the RNA exit channel of the Pol II. These two scenarios were described in several works aimed at identifying additional sequences influencing Pol II termination. The first scenario seems to correlate to GC-rich sequences downstream at the PAS site where Pol II pause gives more time for the exonuclease to reach Pol II (Plant et al. 2005; Gromak et al. 2006). These elements are associated with RNA:DNA hybrids and their resolution by specific helicases (Sen1 in yeast and senataxin in mammals) is essential for termination to occur (Skourti-Stathaki et al. 2011). The second case is more elusive and has been recently described as valid for 78 gene loci after a genomic approach aimed at assessing the generality of this mechanism (Nojima et al. 2013). In this case an AT-rich sequence 1 to 2 kb downstream of the PAS site is prone to co-transcriptional cleavage (Dye & Proudfoot 2001). This element is called co-transcriptional cleavage (CoTC) element and might occur when PAS-proximal pause sites are absent. These AT-rich sequences are highly unstable and susceptible to degradation so that, as soon as they are synthesised, they are attacked by exonucleases allowing fast and efficient termination (West et al. 2008). CoTC cleavage drives the Torpedo mechanism then to occur further downstream the PAS site and the release of the Pol II precedes cleavage and 3' end processing.

1.4 The chromatin template

Eukaryotic DNA is packed around a sophisticated complex of proteins called histones. Two copies of histones H2A, H2B, H3 and H4 combine in an octamer called a nucleosome (Kornberg 1977). The entire DNA is assembled around an array of nucleosomes that take contact with every 146 bp of DNA. The

DNA/nucleosome array is in turn folded into chromatin fibres compacting the long DNA into a few microns within a cell's nuclei.

Although chromatin is a major challenge to overcome for both TFs and Pol II, it provides opportunities for gene regulation. The packaging of the DNA indeed affects genes regulation and it seems to contribute to the setup and maintenance of cell identities (T. Chen & Dent 2014).

Multiple mechanisms contribute to control chromatin assembly and compaction including DNA modifications (*e.g.* cytosine methylation and cytosine hydroxymethylation), histones PTMs, incorporation of histone variants (*e.g.* H2A.Z and H3.3) and ATP-dependent chromatin remodelling. A combination of these modifications could contribute to the regulation of DNA accessibility. Beside the regulation of the chromatin architecture and nucleosome density each histone PTM and each histone variant can constitute a possible docking site for regulatory machines (see below).

More than half a century ago the word 'epigenetics' was coined by Waddington, defining the mechanisms that temporally and spatially control gene activity during development (Waddington 1959; Holliday 2007). Nowadays the word 'epigenetics' is extensively used to group all the DNA sequence-independent mechanisms controlling gene expression, including DNA methylation states and histone modifications (Bernstein et al. 2007; Bonasio et al. 2010). It is important to point out that the instruction provided by the 'epigenome' is functional if considered in the context of the genetic information coded within the DNA.

Dynamic and reversible changes occurring at the chromatin structure are of particular interest in gene activation while long-term modifications such as DNA methylation are mostly associated with genomic imprinting, X-chromosome inactivation, suppression of repetitive elements and long-term gene silencing. DNA methylation occurs in the context of CpG dinucleotide (Jones 2012). CpG regions are mostly present at active promoters and methylation of cytosines by DNMT enzymes causes stable gene repression. On the other hand histone modifications, exchange and repositioning are fast changes correlating with different 'chromatin states'. Several works contributed to elucidating how these chromatin states or signatures are influencing the transcriptional outcome.

1.4.1 Histone post-translational modification

Histone proteins have an amino-terminal tail protruding from the nucleosome core and can be subject to several PTMs (Tan et al. 2011; Tian et al. 2012). Due to the number of combinations the total number of 'chromatin states' is likely inestimable (Cieślik & Bekiranov 2014). Several studies aimed at mapping PTMs described a certain correlation between specific combinations of PTMs and precise functions or genomic loci, reducing the number of possibilities. Histone PTMs include: acetylation, methylation and phosphorylation. Further modifications such as deamination, ADP ribosylation, ubiquitylation, sumoylation and proline isomerisation could also occur but are not going to be introduced here. For a recent review on histone PTMs see Bannister & Kouzarides 2011.

Histone acetylation: The acetylation of lysine residues within histone N-terminal tails is a highly dynamic process under the control of the opposing enzymes; histone acetyltransferases (HATs) and histone deacetylases (HDACs). The addition of an acetyl group to lysine causes neutralisation of the lysine's positive charge, weakening the interaction between histones and DNA (Bannister & Kouzarides 2011). At the same time acetylated lysines can appear as functional docking sites for proteins harbouring specific domains (for a review see Patel & Zhanxin Wang 2013). HATs can be grouped depending on their catalytic domain (see Table 1.4). HATs reside in multi-protein complexes where the specific composition dictates the unique features of each HAT complex. Various chromatin-binding domains can determine the location of HATs complexes docking onto modified histones. HATs are therefore taking advantage of specialised subunits that are able to determine their genomic localisation (Sterner & Berger 2000).

HDACs have opposing actions, restoring the positive charge of the lysine, consistent with their function in transcription repression. In mammals it is possible to list 18 different HDACs divided into four classes, each showing different structure, enzymatic function, subcellular localisation and expression patterns (Haberland et al. 2009; de Ruijter et al. 2003). HDAC activity is controlled through the association with co-repressor and accessory proteins assembled into multi-subunit complexes (Adcock 2006).

Histone phosphorylation: Like acetylation, phosphorylation of histones on serines, threonines and tyrosine dramatically alters their overall charge. Histone phosphorylation has the most disparate functions spanning from regulation of chromatin compaction in meiosis and mitosis to DNA damage repair and transcription regulation (Rossetto et al. 2012). Chromatin compaction and DNA damage response involve extended phosphorylation over several kilobases. However phosphorylation of histones in association with transcriptional regulation resides in defined and restricted genomic loci (Barratt et al. 1994). The most studied phospho-modifications lie within the N-terminal tail of the histone H3, in particular H3S10 and H3S28 (Sawicka & Seiser 2012). Kinase dependent histone modifications were always seen as possible starting points required to induce defined activating signature at promoters. The identification of the kinases involved in transcriptional activation through histone phosphorylation has always been a challenge. Uncoupling the causal effect of activated kinases and the coinciding histone modification is difficult to rule out. Several indirect effects could be observed as in most cases the primary readout is the activation of target genes rather than the appearance of the modification. *In vitro* studies often try to validate the in vivo observations but are often subject to unspecific effects. Table 1.5 presents the kinases that were shown to phosphorylate histone residues *in vitro* or through correlation studies *in vivo*. The best-described kinases are the Aurora family of kinases. Their role is best described in mitosis and meiosis but recent evidence described a possible function in controlling gene activity in quiescent lymphocytes directly associated with target promoters (Frangini et al. 2013).

Beside the identification of the kinase responsible for histone modifications, the consequence of having a phosphate residue on the histone H3 tail is of central interest. Several studies tried to rule out the relation between phosphorylation and other PTMs at the histone tail. The identification of the 14-3-3 factors has been pivotal in understanding the function of signalling induced histone H3 phosphorylation in transcription (Macdonald et al. 2005; Walter et al. 2008). Biochemical studies highlighted how 14-3-3 binding is enhanced when the H3S10 phosphorylation (H3S10P) is accompanied by H3K9 and K14 acetylation (H3K9K14Ac) (Winter et al. 2008). As presented in several studies, 14-3-3 works as a reader of modified histones providing further docking sites for downstream effectors (Drobic et al. 2010; Zippo et al. 2009; K. D. Meyer et al. 2008).

It is going to be crucial to understand how histone phosphorylation is directed to specific loci and in particular to promoters of genes downstream of signalling pathways. Furthermore it is going to be important to identify and describe phosphatases that possibly regulate the dynamicity of these marks that, at the moment, remain understudied.

Histone methylation: Methylation occurs on both lysine and arginine of histone side chains. Unlike acetylation and phosphorylation, histone methylation does not affect the charge of the targeted residue. Lysine can be mono- di- or tri-methylated while arginine may be mono- and symmetrically or asymmetrically di-methylated (for review see Shilatifard 2006). It is possible to distinguish six major classes of lysine methyltransferase (KMTs) (Table 1.6). KMTs act either as single entities or as part of complexes. All the KMTs posses a so-called SET domain (named after the Drosophila Su(var)3–9, Enhancer of zeste [E(z)], and trithorax [trx]) with the exception of Dot1 which presents an unique domain. Histone methylation has diverse roles in gene activation spanning from positive to negative functions. Several so-called readers are capable of interacting with methylated lysine through specific domains. Two key works employing extensive SILAC-based mass spectrometry, using peptides or reconstituted nucleosomes, described an intricate network made of methylated lysine at the histone tails (Vermeulen et al. 2010; Bartke et al. 2010). These studies clearly demonstrate the correspondence between the known biological function of methyl marks and the biological function of the associated reader. For decades histone methylation was supposed to be a fairly stable modification but the identification of (K)-specific demethylase (KDM) showed how dynamic these marks are. The first demethylase, named lysine-specific demethylase 1 (LSD1), was discovered by Shi and colleagues in 2004 (Yujiang Shi et al. 2004). Since then several other demethylases have been discovered making the methyl mark a highly dynamic histone PTM (for a review see Mosammaparast & Yang Shi 2010).

1.4.2 Exchange and repositioning of histones

The primary function of chromatin remodelers is to dynamically regulate DNA accessibility and to control nucleosome positioning (Clapier & Cairns 2009).

There are four main families of chromatin remodeling enzymes (Table 1.7). All four families use a DNA-dependent ATPase to remodel and translocate along the DNA that, working like an icebreaker, weakens the nucleosome-DNA interactions. The four families can be distinguished on the basis of their specialisation imparted by unique domains and subunits (Table 1.7). Overall remodelers have an affinity for nucleosome beside the ability to recognise specific histone PTMs. Chromatin remodelling machines play a crucial role in several biological processes such as replication, dosage compensation, DNA repair and also gene regulation. Regarding nucleosome positioning and its role in gene activation it is important to consider the antagonising effect of remodelers that organise chromatin and those that disorganise or eject nucleosomes (Clapier & Cairns 2009). Organised chromatin mostly correlates with repressed chromatin as the assembly of nucleosome arrays restricts DNA accessibility. On the other hand disordered chromatin and histone ejection favours gene activation (Lorch et al. 2006). A further strategy to activate expression of target genes is to replace canonical histones with variants that associate with the DNA less stably. This is the case with the SWR1 enzymes which are capable of replacing the canonical dimer H2A-H2B, with the variant H2A.Z-H2B generating a less stable nucleosome at promoters of active genes (Schones et al. 2008; Ruhl et al. 2006; Jin et al. 2009).

1.4.3 Chromatin states

One of the major challenges in studying the chromatin template is the generation of predictive models aiming to relate chromatin signatures to regulatory elements. The ability to unequivocally assign a determined histone signature to cis-regulatory DNA elements or silenced/active genes would provide an immense tool to understand diverse biological outcomes.

This systematic approach was recently pursued by the Bernstein lab. Ernst and collaborators made use of a high-throughput pipeline to systematically map nine chromatin marks within nine cell lines (J. Ernst et al. 2011). They used a series of antibodies recognising marks associated with: active promoters (H3 lysine 4 tri-methylated named H3K4me3, di-methylated named H3K4me2 and the histone variant H2A.Z), enhancers (H3K4me2 and mono-methylatednamed lysine 4 of the histone H3 named H3K4me1), active regulatory elements (H3 lysine 9 acetylated

and K27 acetylated named respectively H3K9ac and H3K27ac), actively transcribed regions (H3 lysine 36 tri-methylated named H3K36me3 and H4 lysine 20 mono-methylated named H4K20me1), associated with Polycomb repression (H3 lysine 27 tri-methylated named H3K27me3) and heterochromatin regions (H3 lysine 9 tri-methylated named H3K9me3). A hidden Markov model was applied to summarize the data into nine annotations, corresponding to the nine cell lines in combination with RNA expression data and Pol II distribution. From this analysis it was possible to discriminate 15 chromatin states (Figure 1.7). The distribution of these chromatin states changes across the different cell lines reflecting lineage-specific gene expression. Furthermore enhancers showed a high cell line specificity being proximal to lineage-specific genes. Using this framework it was also possible to predict cis-regulatory elements and in particular TF binding sites within cell-type-specific enhancers. This study was recently expanded using human tissues, blood cells and stem cells corroborating the view that specification is accompanied by a marked transition from diverse possible states to limited configurations (Zhu et al. 2013).

This innovative approach was integrated thereafter with the combinatorial patterning of chromatin regulators. Using a meso-scale localisation assay Ram and colleagues generated a genome-wide map of 29 chromatin regulators (CRs) (Ram et al. 2011). Several observations came out from this study: genomic loci with the same signature show coherent modules of CRs, composed of both activating and repressing enzymes, suggesting a possible tuning mechanism; specific CRs associate to genes with related functions; within different cells CRs are part of the same module but the distribution of these modules changes. Altogether these conclusions are reminiscent of the organisation of TF networks. Garber and colleagues recently described the importance of TF binding networks and their dynamic behaviour in response to stimuli (Garber et al. 2012). The tight correlation between TF binding, transcription dynamics and chromatin changes highlights the leading role of TFs in seeding the primary causal variable.

1.4.4 Signalling to Chromatin and the histone crosstalk

Having described the most recent advance in interpreting the chromatin signature it become clear that histone modification *per se* cannot represent a code

(J.-S. Lee et al. 2010; Sims & Reinberg 2008). Histone PTMs are no different than PTMs associated within any other protein in the cell. It is instead true that groups of modifications can encode a signature or state that correlates with functional elements or outcomes (E. Smith & Shilatifard 2010). Most intriguingly it is possible to consider histone modification as a part of signalling pathways required to assemble these states. A key question therefore is how the chromatin states are set up. Several studies have tried to understand how defined patterns of chromatin modifications are assembled in response to cell differentiation, tissue morphogenesis and acute ectopic stimulation.

A remarkable example comes from the study of mechanisms involved in maintenance or abrogation of Embryonic Stem cell (ESC) pluripotency. ESCs are characterised by a highly plastic and permissive chromatin that allows them to maintain a wide spectrum of differentiative capacity (Meshorer et al. 2006). ESCs are indeed marked by defined chromatin states that are maintained through a signalling-crosstalk. A well-characterised mechanism to establish silenced chromatin comes from the study of DNA methylation (Z. D. Smith & Meissner 2013). This repressive mechanism is initiated by DNA methyltransferases (DNMT). The establishment of repressive chromatin at these loci is caused by an interplay between DNMTs, histone methyltransferase, HDACs and remodelling complexes (for a review see Z. D. Smith & Meissner 2013). This mechanism in ESCs is controlled by specific TFs. Elegant experiments, targeting SP1 to endogenously methylated loci, demonstrated that TF activities are sufficient to induce demethylation (Macleod et al. 1994; Brandeis et al. 1994).

Long-term chromatin silencing opposes the silencing of lineage-specific genes, later required for ESC differentiation. This group of promoters is characterised by the presence of the polycomb group of proteins (for reviews see Morey & Helin 2010; Voigt et al. 2013). This complex could be recruited at genomic loci through a plethora of different mechanisms including specific cis-regulatory DNA elements, defined histone modifications, TFs and non-coding RNA (ncRNA) (Morey & Helin 2010). Although the way polycomb complexes are specifically recruited still remains unclear, their distribution correlates with the establishment of inactive - or poised - promoters and repressed transcription (Barski et al. 2007; Mikkelsen et al. 2007; J. Ernst et al. 2011). It was observed that the distribution of polycomb protein is complementary to the presence of transcriptional activator and

transcribing Pol II (Ku et al. 2008). This and other studies are now leading to the view that polycomb proteins are excluded from active regions instead of being recruited to target regions.

The establishment of active chromatin states in response to extracellular stimuli yields a notable perspective of how dynamic the histone crosstalk could be. As mentioned earlier several kinases are found associated to promoters in response to stimuli (Frangini et al. 2013; H.-M. Zhang et al. 2008a; Madak-Erdogan et al. 2011). Different transcription factors have been shown to harbour defined docking sites recognised by specific kinases. The location of kinases at gene promoters activates histone crosstalks, which in turn establish a specific chromatin signature. One example comes from studies on active nuclear Erk. In fact, Erk is able to activate the progesterone receptor and the protein kinase MSK1, leading to their specific association to target promoters (Vicent et al. 2006). As a consequence, active Msk1 is able to phosphorylate H3S10, which displaces HP1 from the repressed chromatin and allows recruitment of Pol II. A further example comes from a recent study of stem cell differentiation into neurons where JNK was seen in association with chromatin loci together with NF-Y (Tiwari et al. 2012). As for the other kinases JNK also seems to be able to phosphorylate H3S10.

Phosphorylation of histone H3 allows recruitment of 14-3-3 adaptor proteins which was reported to induce a histone crosstalk at the *Fosl1* gene following serum stimulation (Zippo et al. 2009). At *Fosl1* enhancer PIM1 kinase is responsible for the phosphorylation of H3S10 while MSK1/2 seems to affects its promoter (Zippo et al. 2007; Zippo et al. 2009). In this case, 14-3-3 seems to allow specific recruitment of the acetyltransferase MOF only at the *Fosl1* enhancer. MOF-dependent acetylation of H4K16 was shown to be responsible and sufficient for BRD4/pTEFb recruitment and productive transcription. Although captivating, this study leaves several open questions. How are different kinases targeted at different loci? How does H3S10 phosphorylation induce different crosstalks at the *Fosl1* enhancer and promoter?

1.5 The SRF transcription network

Most transcriptional networks embody several of the regulatory mechanisms described so far. Among them, the Serum Response Factor (SRF) regulates the

expression of growth-related IE, cytoskeletal and muscle-specific genes. In particular, SRF was reported to be essential for the control of growth, differentiation, cell motility and circadian clock oscillations.

SRF was initially discovered through a series of studies based on c-fos gene expression. C-fos is a proto-oncogene, a homolog of the FBJ murine osteosarcoma gene v-fos (Curran et al. 1982). C-fos has always been of high interest in the transcription field due to its rapid RNA synthesis after exposure to growth factors and mitogens (Greenberg & Ziff 1984; Greenberg et al. 1985). Studies of the 5'-flanking region of c-fos identified a cis-element named Serum Response Element (SRE) required for c-fos serum induction with enhancer-like characteristics (Treisman 1985; Deschamps et al. 1985). Soon after, through a series of biochemical studies the isolation of the protein encoding for the factor associated to the c-fos SRE led to the discovery of SRF (Treisman 1986). SRF and its consensus sequence - the SRE - were then found to be associated with several other IE genes and surprisingly muscle-specific genes (Mohun et al. 1987; M. Taylor et al. 1989).

1.5.1 MADS-Box transcription factors and SRF DNA recognition

In vitro functional studies of the SRF-DNA binding activity identify a 90 amino acid domain embedded within the SRF polypeptide, which suffices for normal DNA binding activity (Norman et al. 1988; Schröter et al. 1990). Sequence comparisons show several SRF relative factors, including the two S. Cerevisiae MCM1 and ARG80, involved in extracellular signal-regulated and cell type-specific transcription (Norman et al. 1988; Passmore et al. 1988). The DNA binding domain of SRF, MCM1 and ARG80 all include a conserved 56 amino acid sequence, named MADS-box after its founding members (MCM1-Agamous-Deficiens-SRF) (Shore & Sharrocks 1995). All MADS-box containing factors exhibit a conserved core DNA-binding domain composed of an N-terminal MADS box and a C-terminal extension of approximately 30 amino acids (Shore & Sharrocks 1995). The MADS-box is what determines the DNA specificity of each factor allowing them to bind related, but distinct consensus sequences (Pollock & Treisman 1991; Wynne & Treisman 1992). The C-terminal half however is required for dimerization, and in part for co-regulator interactions (Mueller & Nordheim 1991; Shaw 1992; Primig et al. 1991).

The crystal structure of the SRF homodimer, together with its DNA target sequence provided surprising insight into the MADS-box family of TFs (Pellegrini et al. 1995). The SRF homodimer organises on top of the DNA into a three-layered structure, each one composed of the interaction with the same unit from each monomer. As shown in Figure 1.9 at the base of the SRF-DNA structure lays an antiparallel coiled coil, constituted by two amphipathic α–helices contacting the minor grooves of the underlying DNA. The second layer, consisting of a four-stranded antiparallel β–sheet, sits on top of the antiparallel coiled coil layer. Above the β–sheet layer, the C-terminal unit forms irregular coiled pairs of helices with a pair of α–helices on top. The binding of SRF to the DNA forces the latter to bend by 72°, allowing the basic N-terminal regions and loops to contact the major grooves of each half site. The N-terminal regions contact the minor grooves on the opposite side. The disposition and the specific contacts of the N-terminal region provide rationales for the different sequence specificity of MAD-box containing factors such as SRF, MCM1 and MEF2A (Sharrocks et al. 1993; Nurrish & Treisman 1995). The SRF-DNA structure shows how SRF specifically recognises its consensus sequence, named CArG-box which is conserved across all the SREs studied at that time. The CArG-box is defined by a central A/T-rich sequence that allows the formation of a narrow minor groove while flanking GG bases favour bending into the major groove. Structural and functional analysis of the SRF-DNA binding allows the definition of the SRF consensus sequence as $CC(A/T)_6GG$ (Pellegrini et al. 1995; Pollock & Treisman 1990; Treisman 1985).

Recently the specificity of SRF for the CArG consensus was addressed using genome-wide approaches (Valouev et al. 2008; Sullivan et al. 2011; Esnault et al. 2014). SRF binding *in vivo* occurs at perfect or slightly degenerated CArG sequences (70% of SRF's binding events with 0-1 mismatch and 90% of SRF's binding events with 0-2 mismatchs) and its binding intensity correlates directly to the quality of the CArG consensus (Esnault et al. 2014). In contrast to what has been proposed in the past (Herrera et al. 1989), most SRF binding is inducible and surprisingly, inducible SRF sites exhibit a better match to the CArG-box sequence than constitutive sites (Esnault et al. 2014). As introduced previously the association of SRF to its consensus in the genome most probably relies on nucleosome displacement, provided by co-regulators and TF binding nearby. The quality of the consensus reflects the binding cooperation with co-regulators with no

DNA preferences and TFs that are not directly interacting with SRF. On the other hand degenerated consensus seems more associated with loci where SRF directly associates and interacts with co-regulators harbouring defined DNA recognition features (see later).

1.5.2 SRF transcriptional network: The TCF and MRTF branch

Besides studies focused on SRF-DNA interaction, much effort was devoted during the late 80s to characterise SRF's activities and functions (Treisman 1995b). SRF is a constitutive nuclear TF with a poor transactivation activity (Johansen & Prywes 1993; Hill et al. 1993). Indeed most of SRF's functional features were mapped within or in close proximity to its DNA binding domain and its C-terminal activation domain, fused to the GAL4 DBD, showed weak transactivation functions (Treisman 1995a; Johansen & Prywes 1993). However, SRF was reported to be essential for the expression of several genes and in particular to be required for signal-dependent gene activation (Treisman 1986; Prywes & Roeder 1987).

Shedding light on this problem was the discovery of a novel factor of 62 KDa able to form a ternary complex with SRF at the *c-Fos* promoter (Shaw et al. 1989). Diverse studies led to the isolation of three ETS-related factors (see Table 1.1) named Elk-1 (Rao et al. 1989), SAP-1a (Dalton & Treisman 1992) and NET/SAP-2/Erp (Giovane et al. 1994; Lopez et al. 1994), able to form a ternary complex with SRF and therefore named Ternary Complex Factors (TCF). These factors are constituted by three main domains: an N-terminal EST DNA binding domain (Rao et al. 1989; Dalton & Treisman 1992), a basic B-box required for interaction with SRF (Dalton & Treisman 1992; Janknecht & Nordheim 1992) and a C-terminal transactivation domain (Dalton & Treisman 1992) (Figure 1.10). The TCF activation domain is phosphorylated *via* Erk, which is the ending point of the Ras cascade (Gille et al. 1992; Marais et al. 1992; Janknecht, W. H. Ernst, Pingoud, et al. 1993b). The discovery of the TCFs clearly explained how activation of the MAPK signalling pathway *via* mitogens would target SRF regulated genes. However, it was not sufficient to explain other observations. Serum and polypeptide growth factors or TPA (12-O-Tetradecanoylphorbol-13-acetate) were activating distinct pathways leading to SRF activation (Treisman 1995b). Furthermore activation of SRF dependent genes could occur independently of the TCFs in several cell lines (Hill et

al. 1994; Hill & Treisman 1995a). The involvement of a further signalling pathway was disentangled by studies aimed to characterise this TCF-independent activation. SRE containing promoters were shown to respond to active RhoA, CDC42 and Rac1 (Hill et al. 1995). Furthermore changes in actin dynamics could directly influence genes controlled by SRF in the absence of TCFs (Sotiropoulos et al. 1999; Mack et al. 2001; Posern et al. 2002). The hunt for SRF co-regulators involved in muscle-specific gene activation led to the identification of Myocardin (D. Wang et al. 2001a). Later, the identification of Myocardin homologues potentiating SRF activity provided additional clues (D. Z. Wang et al. 2002). Miralles and colleagues finally demonstrated that Myocardin-related transcription factors (MRTFs) are indeed SRF co-activators, linking RhoA and actin dynamics to SRF-dependent transcription activation (Miralles et al. 2003).

In summary SRF is the endpoint of two signalling pathways (Figure 1.8). The Ras-Erk signalling pathway allows SRF-associated TCF activation and gene induction (see chapter 1.6). On the other hand activation of Rho, and the concomitant remodelling of the actin cytoskeleton, induces MRTF's association with SRF and gene activation (see chapter 1.7).

1.6 Signalling to SRF part I: The RAS-Erk signalling pathway and the TCF co-factors

Mitogen-activated protein kinases (MAPKs) belong to a family of kinases whose function is conserved throughout evolution (Widmann et al. 1999). These kinases link receptor tyrosine kinase activation to effector proteins *via* a protein kinase cascade. MAPKs are able to phosphorylate specific serine and threonine residues on target proteins, regulating diverse cellular activities. MAPKs work like triggers to switch target proteins on or off. MAPK is part of a so-called "phosphorelay" system composed of three sequentially activated kinases that ultimately lead to MAPK phosphorylation and activation (G. L. Johnson & Lapadat 2002).

Multicellular organisms present several MAPK subfamilies including, among others, the p38 kinase family, the c-Jun N-terminal kinase family (JNK or SAPK) and the extracellular signal-regulated kinase (ERK) family.

1.6.1 RAS activation and the phosphorelay to ERKS

ERK1 and ERK2 are widely expressed mitogen-activated protein kinases involved in different cellular processes. ERKs were shown to be involved in the regulation of cell adhesion, cell cycle progression, cell migration, cell survival, differentiation, metabolism, proliferation and activation of transcription (Roskoski 2012a).

The activation of Ras within cells is what leads to the activation of the phosphorelay culminating in ERK nuclear accumulation. Ras is a small GTPase that works as a molecular switch changing its status from inactive to active as it releases GDP and associates with GTP (Marshall 1988). This process of activation is guided by guanine nucleotide exchange factor (GEF). The activation of Ras is the result of ligand-induced activation of protein tyrosine-linked receptors (PTKRs). Binding of growth factors to the PTKR extracellular domain induces dimerization and auto-phosphorylation of the receptor in its intracellular domain (Ullrich & Schlessinger 1990). The cytoplasmic phosphorylated residues are therefore recognised by adaptor proteins harbouring specialised domains such as the Src homology 2 (SH2) domain or the phosphotyrosone binding domain (PTB) (Pawson 2004; Schlessinger 2003). These adaptor proteins (*e.g.* Grb2, Nck, Crk, Shc) bring different factors involved in signal transduction at the plasma membrane. In particular Grb2, using its SH2 domain, brings the GEF Sos at the PTKR in proximity to its target Ras (Schlessinger & Bar-Sagi 1994; Pawson 1995). Sos favours the exchange of GDP with GTP activating Ras at the plasma membrane. Further mechanisms could lead to Sos recruitment at the plasma membrane, including Shc recruitment, PTB adaptor proteins or through interaction with IRS1 or FRS2α (Margolis et al. 1999; X. J. Sun et al. 1993; Kouhara et al. 1997).

Ras can also be activated by other GEFs such as the Ras guanyl nucleotide releasing proteins (RasGRPs) (Ebinu et al. 1998). This GEF harbours a domain able to interact with diacylglycerol (DAG). DAG is well known for its action on PKC (reviewed in Griner & Kazanietz 2007). Increased DAG levels at the plasma membrane, in response to PTKRs or G-protein-coupled receptors (GPCRs), leads to phospholipase C (PLC) stimulation. Activation of PLC leads to the hydrolysis of phospholipid phosphatidylinositol 4,5-bisphosphate (PIP2) in turn generating soluble inositol trisphosphate (IP3) and DAG. Proteins harbouring specialized

domains, such as PKC or, as mentioned, RasGRPs, are activated upon recognition of DAG. Phorbol, a natural diterpenes-derivative obtained from plants of the Euphorbiacaeae family, is a widely used substitute for DAG. The most common phorbol ester is 12-O-tetradecanoylphorbol-13-acetate (TPA), also called phorbol-12-myristate-13-acetate (PMA). TPA is able to induce Ras activation independently of mitogen stimulation (Zheng et al. 2005).

Active Ras at the plasma membrane is known to interact and activate several effector proteins such as PI3K and Raf, stimulating numerous intracellular processes (Howe et al. 1992). Raf is a kinase with a limited spectrum of action. Its activity is restricted to MEK1 and MEK2 that ultimately induce ERK activity (Ray & Sturgill 1988; Roskoski 2012b).

1.6.2 Control of Erk activity and their nuclear targets

Erk activity is under the control of scaffold and anchor proteins. Scaffolds are proteins interacting and bringing together multiple components of the signalling module, working as facilitators, enhancing the propagation of the signal. ERKs bind to several scaffold proteins in the cytoplasm including KRS1/2, IQGAP1, MP1, MORG1, β-arrestine, SEF, MKK1 and paxilin (for review see Roskoski 2012a). Anchor proteins on the other hand bind to single elements of the signalling module, maintaining Erk in the cytoplasm in resting conditions. Upon activation, Erk detaches from these anchors and translocates into the nucleus. Erk kinases were shown to shuttle between the cytoplasm and the nucleus and the relative concentration between the two compartments was proposed to control diverse cellular processes (Zehorai et al. 2010).

The mechanism accounting for Erk nuclear translocation has been a matter of debate for several years. Erk could be accumulated in the nucleus as a result of passive diffusion or active transport (Adachi et al. 1999). Erk does not contain either conventional or non-conventional NLS sequences; therefore several complementary mechanisms were proposed (Zehorai et al. 2010). Recent insight into the mechanism of ERKs nuclear localisation has elucidated an alternative mechanism involving a nuclear translocation signal (NTS) (Zehorai et al. 2010; Plotnikov et al. 2011). Phosphorylation of this sequence allows Erk to interact with importin 7 that mediates Erk translocation into the nucleus *via* the nuclear pore.

Changes in Erk localisation have a profound effect on Erk functions. Recent quantitative proteomic works have shown that a fifth of Erk interaction partners change upon stimulation (Kriegsheim et al. 2009). Most of these interactions are dedicated to controlling Erk spatio-temporal behaviour. Nuclear ERK phosphorylates and activates several TFs including, among others, NF-κB, AP-1, c-Myc and members of the ETS family.

1.6.3 The ETS family of TFs and the TCFs

The ETS family of TFs is a crucial end point of the RAS-RAF-MEK-Erk signalling pathway. The ETS proteins are TFs harbouring a conserved ETS domain responsible for their DNA binding specificity and comprising members of the previously described TCF factors (see Table 1.1). ETS belongs to the helix-turn-helix (HTH) superclass of TFs. HTH domains comprise three α-helices. A first α-helix inserts into the major groove of the DNA, accounting for specific DNA recognition; a second α-helix is necessary in the HTH domain to exert unspecific interaction with the DNA backbone; a third α-helix is present in only a few HTH domains, such as the ETS, and stabilises the entire DNA-binding module. The ETS domain is a variant of the HTH superclass (variant termed winged helix architecture) as the three α-helices are articulated on a four-stranded, antiparallel β-sheet layer (Sharrocks 2001). The consensus sequence recognised by this domain presents a central invariant GGA(A/T) core.

The TCFs are a specialised subfamily of ETS TF, able to directly interact with SRF at specific loci where the ETS DNA binding motif flanks the CArG box. TCFs are a good and well-studied example of modular TFs characterised by multiple domains all contributing to a specific function of the protein (Brent & Ptashne 1985). In addition to the ETS domain, the TCFs harbour a unique hydrophobic motif named B box required for direct interaction with SRF (Figure 1.10) (Dalton & Treisman 1992; Shore & Sharrocks 1994). Recently it has been shown that through direct interaction the TCFs are able to specifically drag SRF at its consensus sequence, stabilising the SRF-DNA association even when the CArG motif is heavily degenerated (Esnault et al. 2014). Hassler and Richmond described the structure of the ternary complex SAP-1a, SRF and DNA by X-ray crystallography (Hassler & Richmond 2001 and reviewed in Sharrocks 2002). As

shown in figure 1.9 the dimeric SRF core and the ETS domain of a single SAP-1a molecule are binding to opposite sides of the DNA with the ETS domain occupying the major groove. The unresolved linker between the ETS domain and the B-box suggests a highly flexible region accounting for the already proposed flexibility in spacing between the CArG box and the ETS binding sequence (Treisman et al. 1992). The B-box arranges into a 3_{10}-helix/β-strand/3_{10}-helix structure, contacting the DNA in its major groove and SRF. Extensive hydrophobic contacts, mediated by residues within the 3_{10}-helix and the β-strand of the B-box, make the β-strand join the β-sheet provided by SRF in an antiparallel fashion. The TCFs are also capable of binding genomic sequences independently from SRF at high affinity ETS sequences, as reported in several studies (Esnault et al. 2014; Hollenhorst et al. 2011).

Several ETS members harbour additional domains able to respond to active ERKs. In particular, one-third of the ETS members encode the PNT (or pointed) domain (for a review see Hollenhorst et al. 2011). The PNT domain acts as a docking module for the active kinase interacting *via* a three-dimensional surface. This interaction allows Erk to mediate phosphorylation of several residues adjacent to the PNT domain in response to mitogens. In contrast, members of the TCF subfamily lack the PNT domain and interact with MAPKs *via* linear peptide sequences (Sharrocks 2002; Buchwalter et al. 2004). In the following section, major insights into the functional role of the activation domain of the TCFs are going to be described.

1.6.4 Transcriptional activation *via* TCFs

The stimulation of the MAPK signalling pathway *via* mitogen leads to SRF-TCF target gene activation (Gille et al. 1995; Hill & Treisman 1995b). The TCFs are the only members of the ETS family harbouring an extended C-terminal domain that upon mitogen activation becomes heavily phosphorylated (Marais et al. 1993). Several consensus S/T-P sites for MAP kinases are scattered throughout this domain (Dalton & Treisman 1992; Hill et al. 1993; Marais et al. 1993). All three TCFs are equipped with two MAPK docking sites. The D box is located upstream of the C domain while the FXFP motif lies within the C domain (Figure 1.10). The D domain and the FXFP motif forms a bi-partite docking module recognised by both

Erk and JNK. The phosphorylation of several S/T-P sites was shown to occur both *in vitro* and *in vivo via* Erk activation (Marais et al. 1993; Zinck et al. 1993; Janknecht, W. H. Ernst, Houthaeve, et al. 1993a). An extended comparative analysis of the three TCFs highlights the conservation between seven major S/T-P sites and two hydrophobic amino acids (residues F378 and W379 in the Elk-1 peptide) (M. A. Price et al. 1995; Cruzalegui et al. 1999). In addition Elk-1 presents two further unique S/T-P motifs. Deletion of the nine S/T-P sites or of the hydrophobic residues abolishes transcription in response to Erk activation. A loss of active transcription is caused by the inhibition of steps beyond Elk-1 DNA binding, as under these conditions the ternary complex remains unaffected. Interestingly, mutation of the F and W residues leaves the Erk-dependent phosphorylation unchanged, suggesting that transcriptional activation *via* Elk-1 cannot reflect a simple addition of negative charges but that the hydrophobic residues might be involved in direct interaction with components of the transcriptional machinery (Cruzalegui et al. 1999).

Elk-1 was shown to constitutively interact, through its C domain, with p300 at target promoters (Janknecht & Nordheim 1996; Q. J. Li 2003). Mitogen-dependent phosphorylation of the Elk-1 enhances p300-Elk-1 association and p300 activity. P300 activity is counteracted by HDAC, stably associated with unphosphorylated Elk-1 (S.-H. Yang et al. 2003). Erk-induced phosphorylation of Elk-1 causes de-repression reversing its sumoylation and disassociation from HDAC (Clayton et al. 2006). Although clear under these circumstances, the generality of this mechanism and how TCFs are interacting with p300 remain to be elucidated. A systematic study of the genetic interaction between the TCFs and the known chromatin modifying enzymes remains to be inspected.

A better-characterised mechanism, concerning the Elk-1 activation domain, involves the specific and direct recruitment of MED23, a subunit of the Mediator Complex (see Chapter 1.3.3). Studies based on the E1A conserved region 3 (CR3) lead to the identification of the Mediator subunit Sur-2 (later re-named MED23 (Bourbon et al. 2004), required for both CR3-dependent and Elk-1-dependent gene activation (Boyer et al. 1999; G. Wang et al. 2001b; Cantin et al. 2003). Compelling genetic studies of the C.elegans homologue Sur-2 showed that this Mediator subunit is required for Ras-dependent vulval development (Singh & Han 1995). This genetic interaction was lead back to Lin-1, a TF harbouring an ETS domain.

Functional studies, using MED23$^{-/-}$ mouse embryonic stem cells (mESc), demonstrated that Med23 is a subunit of the Mediator complex that is specifically required for E1A and Elk1-mitogen-dependent activation (Stevens et al. 2002). Furthermore, direct interaction between Elk-1 and MED23 requires the activation domain of Elk-1 to be phosphorylated (Cantin et al. 2003). Further studies identified additional residues required for this interaction. In particular, the already identified hydrophobic residues are fundamental for Med23 direct interaction (Balamotis et al. 2009). The complete loss of Elk-1 dependent activation in the absence of MED23 suggests that direct recruitment of the Mediator complex is the main mechanism of activation (G. Wang et al. 2005). The mechanism underling the interplay between TAD activation, chromatin changes and transcription remains to be fully dissected and elucidated.

The present thesis addresses these issues using the RAS-TCF branch as a model system.

1.7 Signalling to SRF part II: The Rho-actin signalling pathway and the MRTF co-factors

A major part of this thesis will focus on the regulation of SRF *via* the co-regulars MRTF. As introduced earlier this branch of regulation is governed by the Rho-Actin signalling pathway and by dynamic changes in the G-actin concentration. In this section I am going to outline the main mechanisms leading to Rho activation and changes in actin filamentation. Focus will be dedicated to the nuclear functions of actin and the regulation of the MRTF co-factors within the nucleus.

1.7.1 Globular and filamentous actin regulation

Actin is the most abundant protein in living cells, having an estimated concentration ranging from tens to hundreds of micromoles (Remedios et al. 2003). Furthermore actin is also one of the most conserved proteins across all phyla (Remedios et al. 2003).

Major mechanisms are under the control of this simple globular protein including, among others, muscle contraction, cell motility, cell division, vesicle and organelle movement, cell shape and transcription. All these functions and processes are influenced by the dynamic equilibrium between monomeric globular

actin (G-actin) and polymerized filamentous actin (F-actin) (Pollard & Cooper 2009). Actin filamentation is a tightly controlled process that includes the interplay of several actin-binding proteins (ABP).

The structure of the actin monomer is composed of two α/β lobes or domains of different size representing the “ATPase fold” (for review see Dominguez & Holmes 2011). Small contacts are observed between the two domains allowing the assembly of two separate clefts. The upper cleft is able to bind ATP or ADP and the divalent cation Mg^{2+}, providing further interaction between the two lobes. The lower cleft is instead composed of several hydrophobic residues essential for contacts between actin subunits within the filament and most ABPs. Monomeric actin is an inefficient ATPase while its activity is enhanced several fold once incorporated into filaments. The high ATP concentration in cells, the actin avidity for ATP rather than ADP and the slow ATPase activity of monomeric actin lead to an overall higher concentration of G-actin associated with ATP.

F-actin is a double helical polymer of G-actin molecules arranged head-to-tail (Dominguez & Holmes 2011). This disposition gives a functional polarity to the actin filament defining two distinct extremities: barbed and pointed ends. Overall, due to defined association and disassociation constants, at steady state ATP-actin associates at the barbed end while ADP-actin disassociate from the pointed end leading to a slow treadmilling of actin subunits from the barbed end to the pointed end. This process involves a cycle of “exhausted” ADP-actin and regenerated ATP-actin that respectively associates and disassociates from the growing filament.

The regulation of ATP hydrolysis and ADP to ATP exchange is crucial for the regulation and the speed of treadmilling (Pollard & Borisy 2003). Cells are equipped with several ABPs, falling into more than 60 classes, known to regulate diverse aspects of actin polymerisation. These proteins include, among others, Profilin, thymosin, ADF/Coffilin and gelsolin (for review see Dominguez & Holmes 2011; Bugyi & Carlier 2010). In particular thymosin associates with G-actin in order to buffer the polymerizing process while Profilin favours the exchange of ADP for ATP, promoting the addition of ATP-actin at the barbed end. As mentioned the disassembly of exhausted ADP-actin at the pointed end is also crucial to promote filament growth at the barbed end. Cofilin and gelsolin are notable factors that directly bind actin monomers within the filament producing a cut by changing the actin conformation.

Beside the regulation of actin microfilament growth and treadmilling, the formation of new filaments and branches is also highly controlled. Nucleation of filaments by monomeric G-actin is unfavourable due to the high instability of short actin oligomers (Pollard & Cooper 2009). The initiation and branching of new filaments within cells is favoured by an actin-filament's nucleation process and involves different proteins such as formin, the ARP2/3 complex and several nucleation-promoting factors (NPFs) (for review see Pollard 2007; Bugyi & Carlier 2010). ARP2/3 with its structure mimics a G-actin dimer favouring the nucleation of new filaments. In addition ARP2/3 directly interacts with F-actin, promoting the growth of branches from already existing filaments. Formin is instead involved in the control of filament growth or elongation. Formin catalyses the rapid addition of actin monomers at the barbed end of the filament through profiling-actin recruitment and prevents capping proteins from associating and blocking the filament growth.

All these processes are accurately regulated in response to extracellular and intracellular signals and the Rho-family of GTPases play a crucial role in mediating this intracellular response.

1.7.2 The Rho signalling pathway

Rho GTPases are small monomeric G proteins belonging to the Ras family of G proteins. They comprise 20 members including the most studied RhoA, Rac1 and CDC42 (Heasman & Ridley 2008). As described for Ras, Rho GTPases switch between active GTP-bound forms and an inactive GDP-bound form. The transition between active and inactive is regulated by three sets of proteins. Guanine nucleotide exchange factors (GEFs) promote the release of GDP and association with GTP, leading to activation of the Rho GTPases (Schmidt & A. Hall 2002). GTPase-activating proteins (GAPs) enhance Rho catalytic activity, leading to its inactivation (Tcherkezian & Lamarche-Vane 2007). Finally guanine nucleotide-disassociation inhibitors (GDIs) maintain Rho in an inactive state by inhibiting exchange and preventing their localisation at the membrane (DerMardirossian & Bokoch 2005). Rho GTPases regulate several effector proteins controlling the polymerisation equilibrium of G-actin and F-actin in the cytoplasm. The activation of

Rho GTPases is associated with a net increase in F-actin at the expenses of G-actin.

The activation of Rho GTPases is mediated by receptor tyrosine kinases and G-protein coupled receptors (GPCRs). Several GEFs are induced upon signal stimulation, leading to the activation of the Rho GTPases. Active Rho GTPases imping on several mechanisms affecting actin polymerisation. These include the activation of WASP (Wiskott-Aldrich syndrome protein) that leads to ARP2/3 activation and actin polymerisation, activation of LIM kinase that phosphorylates and represses cofilin and the stimulation of Diaphanous-related formins.

Changes in G- to F-actin ratio have diverse functional outcomes. Beside the mechanisms involved in actin filament assembly, dedicated to motile cellular behaviours, actin itself works as a signalling molecule affecting several effectors and takes part in diverse complexes. In particular, recent works devoted a lot of effort in elucidating the existence and the functional role of actin within the nucleus.

1.7.3 Nuclear actin

The presence of actin in the nucleus has always been a matter of debate (reviewed in Pederson & Aebi 2002). Its low abundance in the nucleus and the difficulty of detecting any stainable filament made it hard to believe in its existence. In 1977 is was reported that actin is in dynamic equilibrium between the cytoplasm and the nucleus of Xenopous oocytes (T. G. Clark & Merriam 1977). Later on microinjection experiments of ABPs and actin antibodies showed that nuclear actin is required for transcription of several genes (Scheer et al. 1984). Actin was also reported to interact directly with all three Polymerases, to be part of chromatin remodelling complexes and to play a crucial role in pre-mRNA processing and export (Percipalle 2013).

To understand all the potential processes controlled by nuclear actin, it is essential to understand its regulation. Recent works have investigated this topic, trying to elucidate how nuclear actin dynamically communicates with the cytoplasm and how it could be potentially regulated by polymerisation.

Dopie and collegues recently showed that cytoplasmic and nuclear actin are in dynamic exchange (Dopie et al. 2012). The characterisation of nuclear actin exports and imports using both fluorescence loss in photobleaching (FLIP) and

fluorescence recovery after photobleaching (FRAP) experiments suggest a three-step mechanism for actin import and retention. In particular, most nuclear G-actin is quickly exported from the nucleus *via* exportin-6. The remaining nuclear actin is likely to be associated with transcriptional machinery as nuclear filaments are highly unstable. On the other hand monomeric G-actin import relies on importin 9 and cofilin. Analysis of the import kinetic showed that the stabilisation of the imported actin could be initially promoted by short, highly unstable, polymeric actin pools and subsequently by its incorporation into stable nuclear complexes.

The polymerisation of nuclear actin could also affect its functions. Several regulators of actin polymerisation and turnover, including formins, have been detected in the nucleus. Indeed actin can form highly dynamic and submicron long filaments within the nucleus as shown by McDonald and colleagues (McDonald et al. 2006). Furthermore dynamic exchange of G- to F-actin within the nucleus is subject to signal-induced regulation of formin proteins (Baarlink et al. 2013). The activation of actin filamentation within the nucleus leads to a net reduction of monomeric G-actin with obvious functional effects including activation of the MRTF co-factors (see later). A further mechanism controlling actin polymerisation in the nucleus was described recently and involves the so-called Molecules Interacting with CasL (MICAL) family of flavoprotein monooxygenases (for a review see Giridharan & Caplan 2014). MICAL proteins use their redox potential to oxydize actin, causing disassembly of actin filaments. In particular, MICAL-2 is enriched in cell nuclei and its activity leads to reduced monomeric nuclear G-actin (Lundquist et al. 2014). This observation seems contradictory as MICAL-2 induces F-actin depolymerisation within the nuclear compartment and the mechanism leading to a possible export of the oxidized actin is not yet understood. A further intriguing study suggests a potential role for the nuclear envelope protein emerin in modulating nuclear actin polymerisation (C. Y. Ho et al. 2013). Although more studies are needed to elucidate the role and regulation of actin within the nucleus, it starts to become clear that cells have developed diverse mechanisms to control its behaviours. Furthermore changes in nuclear actin behaviours are not a simple consequence of cytoplasmic actin dynamics but they posses a dedicated set of tools and modes of regulation.

1.7.4 The Myocardin family of SRF co-factors

As introduced earlier, activation of SRF target genes can occur independently of the TFCs in response to actin dynamics. The direct link between changes in actin behaviour and the activation of SRF target genes was disentangled with the discovery of the Myocardin family of transcription factors (Pipes et al. 2006). The discovery that MRTFs are under the direct control of Rho GTPases and actin dynamics offered the first evidence of a previously unknown link between the actin cytoskeleton and gene expression.

Myocardin is the founding member of this family of factors, first identified in a bioinformatic screen for cardiac-associated genes (D. Wang et al. 2001a). Although named as TFs, Myocardin and members of its family are unable to bind DNA directly and their activity relies on SRF (D. Z. Wang et al. 2002). Myocardin was found to be able to stimulate transcription of CArG-box associated genes and it can form a stable ternary complex with SRF (D. Z. Wang et al. 2002). Myocardin expression is mainly restricted to the cardiovascular system and in subsets of smooth muscle cells (SMCs). Expressed in a broader range of cells and tissues are the Myocardin-related factors MRTF-A and MRTF-B (Mercher et al. 2001; D. Z. Wang et al. 2002; Ma et al. 2001). MRTF-A and -B are co-factors dedicated to SRF while Myocardin was also found associated to MEF2 (Wu et al. 2010; Creemers et al. 2006).

Myocardin family members share homology in several functional domains (Figure 1.11). Focus will be dedicated to the description of major domains within the MRTF co-factor proteins.

At the N-terminal part of MRTF proteins, three RPEL motifs are organised in a domain required for direct G-actin interaction (Miralles et al. 2003; Guettler et al. 2008; Mouilleron et al. 2008). The RPEL motif is composed of an RPxxxEL core sequence. The three RPEL motifs in the MRTF co-factors contain this conserved core sequence while Myocardin shows divergent RPEL1 and 2, devoiding it of actin regulation (Guettler et al. 2008). Recently a new role for Myocardin's RPELs has been proposed, contacting Arps proteins (see section 1.7.5; Morita & Hayashi 2014). Each RPEL motif of MRTF is assembled in an L-shaped structure composed of two α-helices contacting the hydrophobic lower cleft of monomeric actin (Mouilleron et al. 2008). The R residues are essential for MRTF-actin

interaction as MRTF mutants with R to A substitutions show reduced actin affinity (Guettler et al. 2008). In the context of the three RPEL motifs the whole domain is able to interact with up to five actins (Mouilleron et al. 2011). Cooperative binding occurs between the three RPELs allowing a primary three-actin structure. Linker sequences between RPEL1-2 and RPEL2-3 provide further interaction for two additional actins with low affinity (Mouilleron et al. 2011). The information provided by different crystal structures and the functional dissection of this domain suggests that the interaction between each motif and actin does not occur independently from each other. Cooperative association between the actin molecules associated to the same RPEL domains makes the MRTF-actin interaction unique (Guettler et al. 2008).

Embedded within the REPL motifs is an extended bipartite nuclear localisation signal, composed of two basic elements B3 and B2, which is required for import of MRTF proteins (Pawłowski et al. 2010). The extended NLS sequence interacts with importin α and β. The basic elements are also conserved in Myocardin and its dependency on importin β was reported in the lab (MK Vartiainen, unpublished data). Importin proteins are going to be in direct competition with actin at MRTFs' N-terminal domain. Oppositely, at the Myocardin N-teminal domain permissive interaction of importins will result in Myocardin constitutive nuclear localisation. However in the context of MRTF mutants, without their N-teminal domain, deletion of a further basic region, B1, downstream of the RPEL motif accounts for MRTF nuclear localisation (Miralles et al. 2003; Zaromytidou et al. 2006).

The B1 box is an element primarily required for ternary complex formation (Miralles et al. 2003). Several hydrophobic and positively charged residues within the C-teminal half of the B1-box are crucial for the association of SRF with the target DNA (Zaromytidou et al. 2006). Residues, within this element, required for MRTF nuclear localisation can be separated from the one required for ternary complex formation. An additional hydrophobic region, named Q-box as it is rich in Q amino acids, provides further favourable interactions with SRF (Zaromytidou et al. 2006). Although not critical for MRTF activity, the Q-box was also found to affect MRTF nuclear localisation (Miralles et al. 2003).

Several studies have shown that the TCFs and MRTFs are competing for SRF, suggesting that they interact with overlapping surfaces on SRF (Miralles et al.

2003; Murai & Treisman 2002; Zhigao Wang et al. 2004). The detailed characterisation of the ternary complex formation proposed that MRTF and Elk-1 interact differently on the same surface (Zaromytidou et al. 2006). Footprint analyses have shown that MRTF also interacts with sequences of DNA proximal to the CArG-box, and the SRF-dependent DNA bending facilitates MRTF association (Zaromytidou et al. 2006). Genome-wide analysis showed AP-1 and TEAD associated cis-regulatory elements are frequently enriched at MRTF-specific SRF sites (Esnault et al. 2014). Direct, unspecific contacts between MRTF and these DNA elements are conceivable, although is not possible to exclude interplay between MRTF and these TFs.

Moving further towards the C-terminal domain of MRTF it is possible to identify a 35 amino acid domain named SAP (SAF-A/B, Acinus, Pias). This domain is predicted to adopt two amphipathic helices with similar topology to several DNA binding proteins (Aravind & Koonin 2000). The SAP domain is conserved in all Myocardin-family members and deletion from the Myocardin protein shows impairment of a defined set of genes (D. Wang et al. 2001a). No significant reduction in gene expression and target association was observed for MRTF proteins lacking this domain (Miralles et al. 2003; Zaromytidou et al. 2006). In light of recent findings, analysis of the contribution of this domain to cooperative binding between MRTF and SRF on target DNA would provide further insights.

Myocardin-family proteins are equipped with a leucine-zipper (LZ) like domain downstream of the SAP domain (Figure 1.11). The LZ motif mediates homo- and heter-dimerization both in MRTF-A, MRTF-B and Myocardin (Miralles et al. 2003; Selvaraj & Prywes 2003; Zaromytidou et al. 2006). Myocardin homo-dimerization appears to be weaker when compared to MRTF, and does not occur in the context of ternary complex formation with SRF (Miralles et al. 2003; Zaromytidou et al. 2006; Zhigao Wang et al. 2003). However, full transcriptional activity of Myocardin requires *in vivo* a functional LZ (Zhigao Wang et al. 2003; Du et al. 2004).

The C-terminal regions of Myocardin and the MRTFs are more divergent and work as a trans-activation domain (Miralles et al. 2003; D. Wang et al. 2001a). When fused to the LexA-DNA binding domain they activate a LexA reporter gene (Miralles et al. 2003). Our recent genome-wide studies have shown that active nuclear MRTF induces gene expression through different mechanisms (Esnault et

al. 2014). On one hand, active nuclear MRTFs are able to induce the recruitment of Pol II and the subsequent modification of its CTD in response to the Rho-actin pathway. On the other hand several SRF-MRTF targets still show a serum-dependent recruitment of Pol II in the absence of nuclear MRTFs. At these target genes the modification of the Pol II CTD is under the direct control of nuclear MRTF and is required for gene stimulation (Esnault et al. 2014). The dissection of the MRTF activation domain, which is heavily phosphorylated upon nuclear localisation, would provide further insight into this mechanism.

1.7.5 Actin-mediated regulation of MRTF

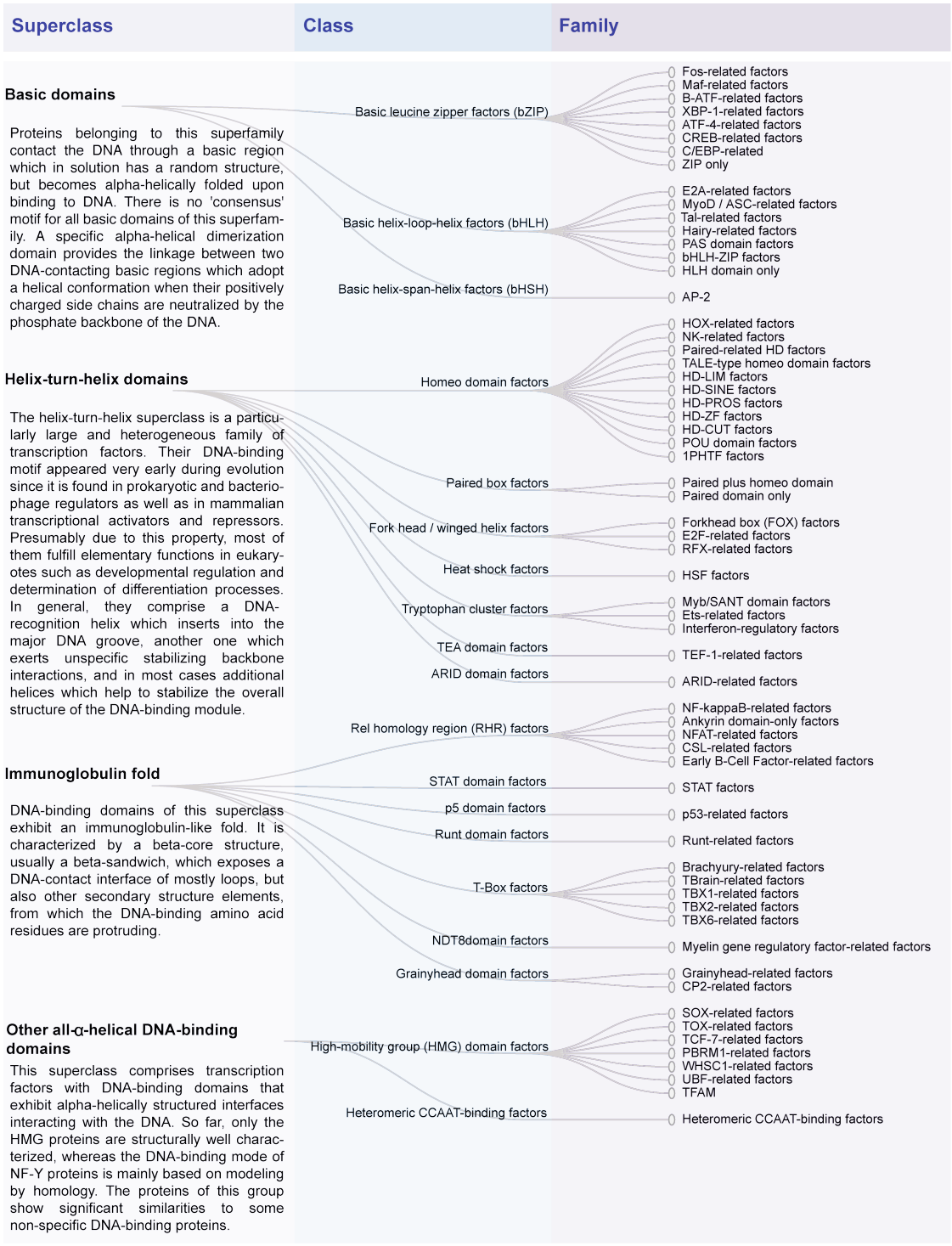
MRTFs co-regulators are under the direct control of G-actin. Changes in actin dynamics have a drastic effect on MRTF functions. The regulation of nuclear import and export is the most well known mechanism under the control of G-actin. In low actin polymerisation states, MRTFs are found mainly in the cytoplasm (Miralles et al. 2003). The overall distribution of MRTF under resting conditions is caused by an actin-dependent promotion of MRTF nuclear export *via* Crm1 (Vartiainen et al. 2007). MRTFs are constantly shuttling and blockage of Crm1, using LMB, allows MRTF nuclear accumulation leaving their interaction with actin unaffected (Vartiainen et al. 2007). Indeed induction of actin filamentation, *via* Rho stimulation, does not affects MRTF import rate while it reduces its rate of export (Vartiainen et al. 2007). However, exogenous overexpression of actin affects MRTF nuclear import, probably due to an increased occupation and occlusion of the B2 and B3 boxes (see section 1.7.4; Pawłowski et al. 2010; Vartiainen et al. 2007). MRTF nuclear export can be accelerated using de-polymerizing agents such as Latrunculin B (LatB), which sequesters actin monomers, increasing the overall concentration of cellular G-actin (Vartiainen et al. 2007).

MRTF can be specifically accumulated in the nucleus using Cytochalasin D (CD) (Miralles et al. 2003; Sotiropoulos et al. 1999). CD is an actin depolymerizing drug known to interact with the barbed end of actin filaments, inhibiting both the association and disassociation of actin monomers, it promotes ATP hydrolysis and induces dimerization of G-actin (Dominguez & Holmes 2011). Opposing LatB, CD directly competes with the RPELs on actin promoting MRTF release, nuclear accumulation and gene activation. The transcriptional response elicited by CD is

substantially different if compared to the Rho-dependent stimulation of MRTF (Esnault et al. 2014). CD can be envisioned as the minimal requirement to induce MRTF activity.

The mechanism leading to MRTF nuclear accumulation is well studied and characterised, while its regulation once in the nucleus remains elusive. Recent studies have shown a new mechanism of regulation regarding the founding member Myocardin (Morita & Hayashi 2014). As already described, Myocardin does not interact efficiently with actin due to a degeneration of the RPEL sequences (Guettler et al. 2008). Morita and colleagues described an Arp5-dependent repression of Myocardin transcriptional activity (Morita & Hayashi 2014). Arp5 is an actin related protein that was reported to directly interact with Myocardin *via* its degenerated RPEL motifs and to the SRF DBD. Furthermore Arp5 seems to affect ternary complex formation, directly impairing both SRF-DNA and Myocardin-SRF interaction (Morita & Hayashi 2014). This mechanism of regulation seems dedicated to Myocardin as MRTFs are unable to interact with Arp proteins and MRTF activity seems unaffected. Preliminary evidence collected in the lab has shown that MRTF activity is negatively controlled by nuclear actin (Vartiainen et al. 2007). Passive accumulation of MRTF into the nucleus *via* LMB does not promote target gene activation. Under this circumstance MRTF-DNA association with target promoters seems unaffected while its transactivation activity requires its disassociation from nuclear actin (Vartiainen et al. 2007).

This thesis will address these issues in detail.


Superclass
Class
Family
Basic domains
Proteins belonging to this superfamily contact the DNA through a basic region which in solution has a random structure, but becomes alpha-helically folded upon binding to DNA. There is no 'consensus' motif for all basic domains of this superfamily. A specific alpha-helical dimerization domain provides the linkage between two DNA-contacting basic regions which adopt a helical conformation when their positively charged side chains are neutralized by the phosphate backbone of the DNA.
Basic leucine zipper factors (bZIP)
Fos-related factors
Maf-related factors
B-ATF-related factors
XBP-1-related factors
ATF-4-related factors
CREB-related factors
C/EBP-related
ZIP only
Basic helix-loop-helix factors (bHLH)
E2A-related factors
MyoD / ASC-related factors
Tal-related factors
Hairy-related factors
PAS domain factors
bHLH-ZIP factors
HLH domain only
Basic helix-span-helix factors (bHSH)
AP-2
Helix-turn-helix domains
The helix-turn-helix superclass is a particularly large and heterogeneous family of transcription factors. Their DNA-binding motif appeared very early during evolution since it is found in prokaryotic and bacteriophage regulators as well as in mammalian transcriptional activators and repressors. Presumably due to this property, most of them fulfill elementary functions in eukaryotes such as developmental regulation and determination of differentiation processes. In general, they comprise a DNA-recognition helix which inserts into the major DNA groove, another one which exerts unspecific stabilizing backbone interactions, and in most cases additional helices which help to stabilize the overall structure of the DNA-binding module.
Homeo domain factors
HOX-related factors
NK-related factors
Paired-related HD factors
TALE-type homeo domain factors
HD-LIM factors
HD-SINE factors
HD-PROS factors
HD-ZF factors
HD-CUT factors
POU domain factors
1PHTF factors
Paired box factors
Paired plus homeo domain
Paired domain only
Fork head / winged helix factors
Forkhead box (FOX) factors
E2F-related factors
RFX-related factors
Heat shock factors
HSF factors
Tryptophan cluster factors
Myb/SANT domain factors
Ets-related factors
Interferon-regulatory factors
TEA domain factors
TEF-1-related factors
ARID domain factors
ARID-related factors
Rel homology region (RHR) factors
NF-kappaB-related factors
Ankyrin domain-only factors
NFAT-related factors
CSL-related factors
Early B-Cell Factor-related factors
Immunoglobulin fold
DNA-binding domains of this superclass exhibit an immunoglobulin-like fold. It is characterized by a beta-core structure, usually a beta-sandwich, which exposes a DNA-contact interface of mostly loops, but also other secondary structure elements, from which the DNA-binding amino acid residues are protruding.
STAT domain factors
STAT factors
p5 domain factors
p53-related factors
Runt domain factors
Runt-related factors
T-Box factors
Brachyury-related factors
TBrain-related factors
TBX1-related factors
TBX2-related factors
TBX6-related factors
NDT8domain factors
Myelin gene regulatory factor-related factors
Grainyhead domain factors
Grainyhead-related factors
CP2-related factors
Other all-α-helical DNA-binding domains
This superclass comprises transcription factors with DNA-binding domains that exhibit alpha-helically structured interfaces interacting with the DNA. So far, only the HMG proteins are structurally well characterized, whereas the DNA-binding mode of NF-Y proteins is mainly based on modeling by homology. The proteins of this group show significant similarities to some non-specific DNA-binding proteins.
High-mobility group (HMG) domain factors
SOX-related factors
TOX-related factors
TCF-7-related factors
PBRM1-related factors
WHSC1-related factors
UBF-related factors
TFAM
Heteromeric CCAAT-binding factors
Heteromeric CCAAT-binding factors

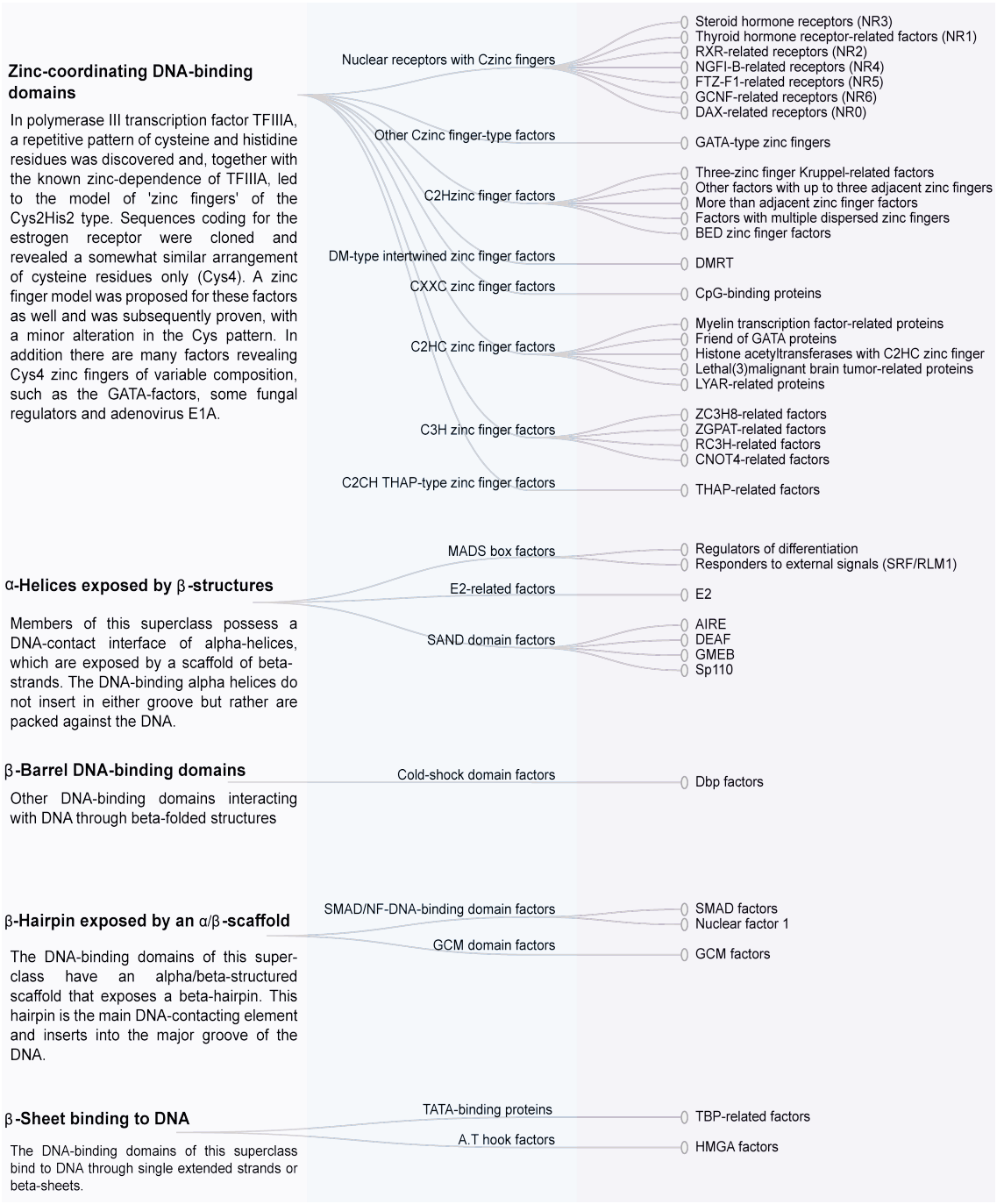


Table 1.1 Transcription factors family.

The classification is based on the TFClass database (Wingender et al. 2013). This classification comprises five levels (superclass, classes, families, genera and factor species). In this table I have listed the first three levels of the database.

O-linked modifications

Phosphorylaton: Addition of a phosphate group through an ester bond to polar amino acids results in increased negative charge in the vicinity of the modification.

Glycosylation: Addition of an O-linked N-acetyl-glucosamine monosaccharide to polar amino acids through a β-linkage. Competitive with phosphorylation, glycosylation does not alter the charge of the protein.

N-linked modifications

Methylation: Addition of CH3 to basic amino acids results in increased hydrophobicity. Unlike acetylation, mono-, di, and tri-methylation of lysine is not charge neutralizing but increases the effective radius of the positive charge by replacing hydrogen with bulky methyl groups.

Acetylation: Acetylation of lysine groups is charge neutralizing and competitive with ubiquitination and sumoylation.

Sumoylation: Addition of one or many ~100 amino acid peptides through a highly labile ε-amino isopeptide bond to lysine results in greatly increased protein bulk. Poly- sumoylationcan occur via lysines within the SUMO moiety.

Ubiquitination: Addition of one or many ~76 amino acid peptides through an ε-amino isopeptide bond to lysine or through a peptide bond to the amino terminus. Poly-ubiquitination can occur through different lysines in the ubiquitin peptide to form a variety of chains.

Table 1.2 TFs post-translational modifications.

Adapted from Filtz et al. 2014.

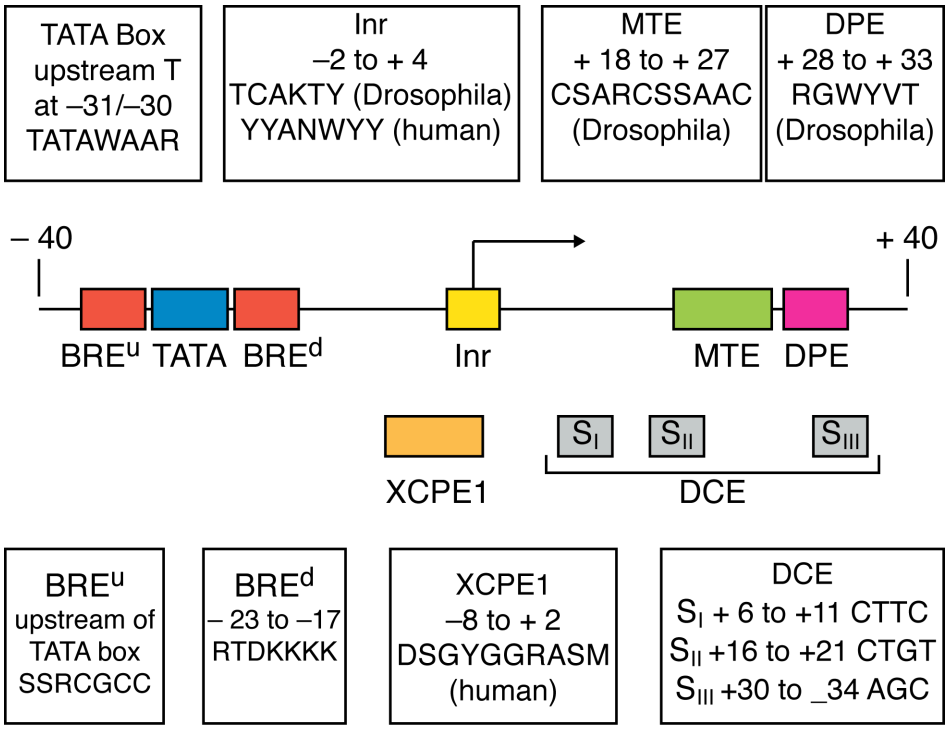


Figure 1.1 Core promoter elements.

Schematic representation of the main elements found at core promoters (adapted from Juven-Gershon & Kadonaga 2010).

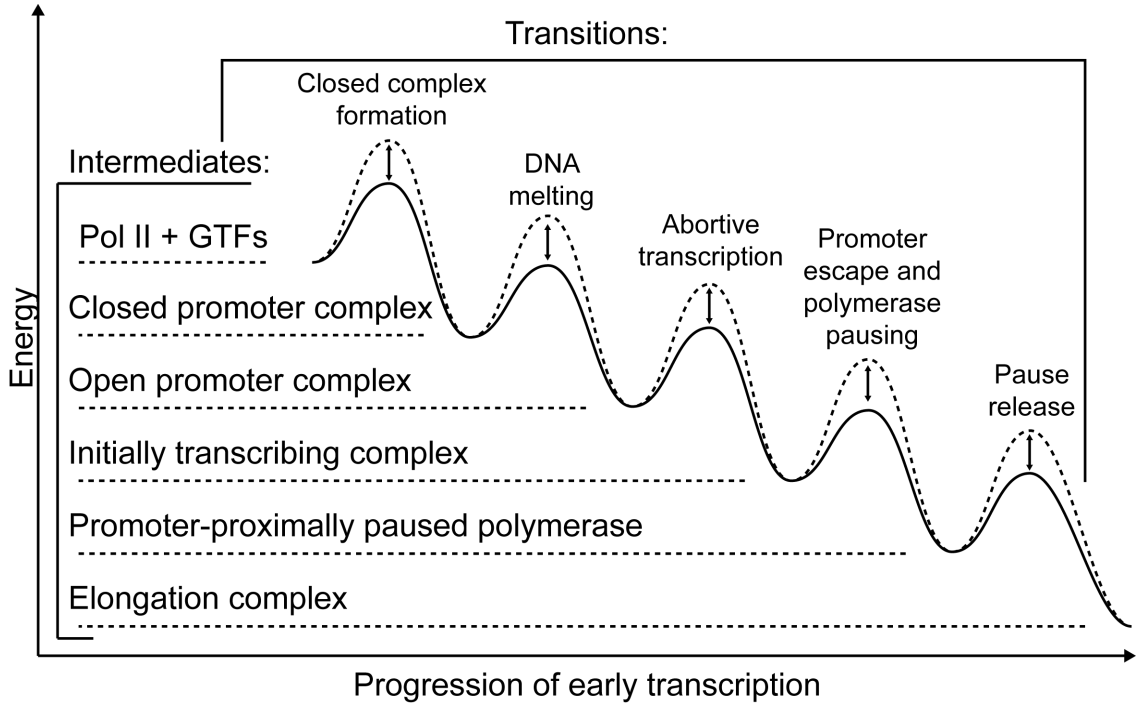
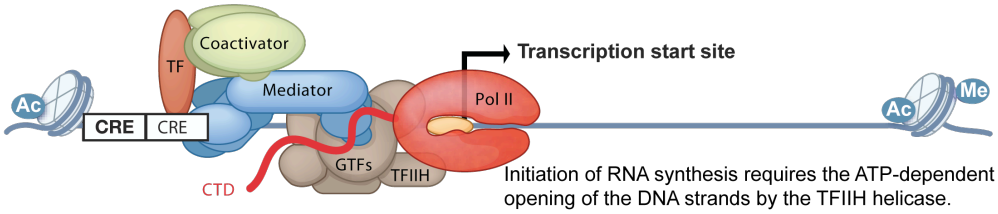


Figure 1.2 Major transitions during early transcription.

Early transcription can be divided into five major transitions. The graph shows the progression of the transcriptional process on the x axis while the y axis depicts the energy barrier that need to be overcome to transit from one condition to another. The energy barrier between each state can be increased or decreased by regulators (double-headed arrows). (Adapted from Michel & Cramer 2013).

Assembly of the preinitiation complex (PIC) and initiation

TFs associate with cis-regulatory elements (CRE) to promote assembly of the PIC.
TFs can stimulate PIC formation directly through interaction with the Mediator complex and GTFs
TFs also recruit coactivators that can remove nucleosomes from the promoter or modify histones proteins, for example by acetylation (Ac) or methylation (Me).



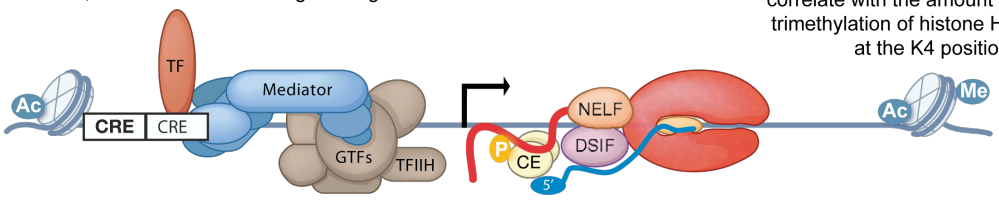
Initiation of RNA synthesis requires the ATP-dependent opening of the DNA strands by the TFIIH helicase.

The Pol II C-terminal domain (CTD) is comprised of many repeats of the sequence YSPTSPS that undergo posttranslational modifications during the transcription cycle but are generally unmodified in the PIC.

Pausing during early elongation

After a 20–60 nt RNA has been synthesized, pause-inducing factors DSIF and NELF bind the early elongation complex, inhibiting further transcription. The duration of pausing depends on the rate of recruitment of factors that trigger pause release, which is variable from gene to gene and under different cell conditions.

Levels of promoter Pol II often correlate with the amount of trimethylation of histone H3 at the K4 position.



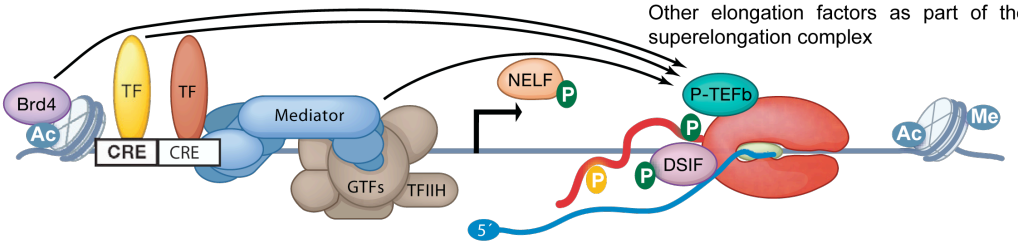
The GTF component TFIIH contains the kinase Cdk7, which phosphorylates the Pol II CTD at the S5 and S7 residues.

The capping enzyme complex (CE) adds a protective 5′ cap to the nascent RNA. CE recruitment is stimulated by its interactions with DSIF and S5-phosphorylated CTD.

Pause release

Release of paused Pol II is triggered by the kinase P-TEFb, which phosphorylates pause-inducing factors to dissociate NELF and promote productive elongation.

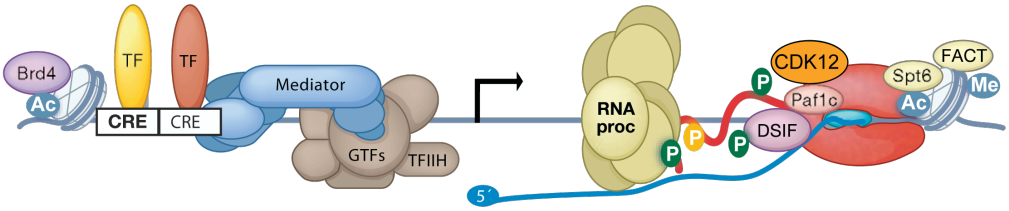
Modes of P-TEFb Recruitment:
TFs (e.g., MYC, NF-κB)
Mediator (Med26 subunit or Cdk8 module)
Acetylated histones and Brd4
Other elongation factors as part of the superelongation complex



Productive elongation

Following phosphorylation by P-TEFb, DSIF helps recruit factors that stimulate productive elongation. These factors (e.g., Paf1c) enhance the rate or processivity of RNA synthesis.

Transcription through nucleosomes and their efficient reassembly in the wake of transcribing Pol II are facilitated by histone chaperones like FACT and Spt6.



The Pol II CTD is also phosphorylated at S2 by the kinase Cdk12/cyclinK, creating a binding platform for factors that aid in RNA processing and transcription through chromatin.

Figure 1.3 Transcription regulation.

Step by step description of the mechanisms involved in transcriptional regulation. Emphasis is placed on the regulation of transcriptional elongation through promoter-proximal pausing of Pol II. Steps are outlined starting with the assembly of PIC at gene promoters to productive synthesis of RNA (adapted from Fromm et al. 2013).

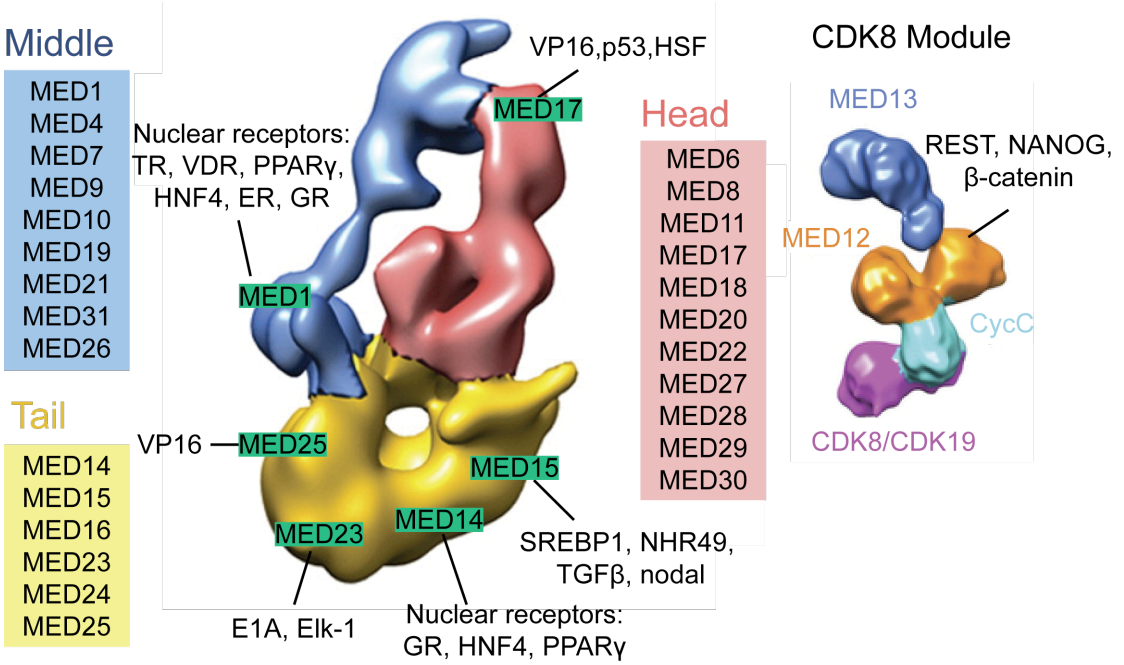


Figure 1.4 Modular structure of Mediator and interactions with diverse factors.

The recently published EM structures of the human Mediator complex and CDK8 module are shown (Tsai et al. 2013; Tsai et al. 2014). On the left hand side of the figure the EM structure of the main Mediator complex is coloured according to the main modules: Head (red), middle (blue) and tail (yellow). On the right hand side the CDK8 module (CKM) subunit organisation is shown. known TF interactions are highlighted. ER, oestrogen receptor; GR, glucocorticoid receptor; HNF4, hepatocyte nuclear factor; NHR49, nuclear hormone receptor 49; PPARγ, peroxisome proliferator-activated receptor-γ; SREBP1, sterol regulatory element-binding protein 1; TGFβ, transforming growth factor-β; VDR, vitamin D3 receptor (Malik & Roeder 2010).

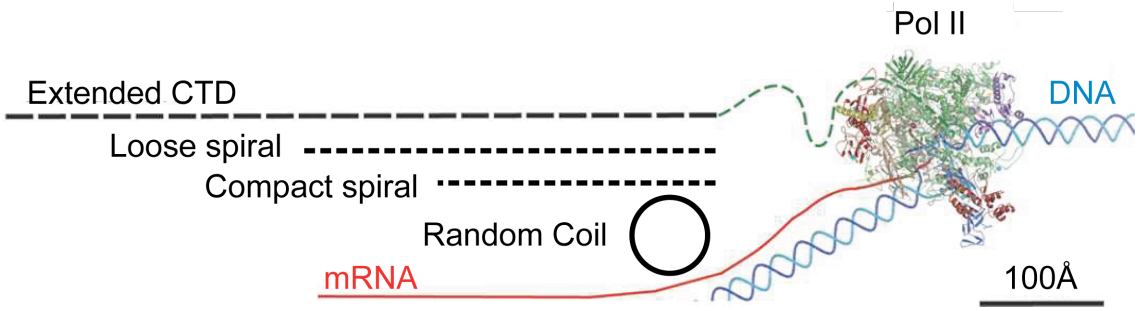


Figure 1.5 Elongating Pol II and size of the CTD.

On the right the 12 subunit of the elongating yeast Pol II is shown as a ribbon diagram with each subunit in a different colour. The DNA is shown schematically in blue while the nascent RNA in red. The relative size of the Pol II largest subunit linker is shown in green while the extended CTD is schematically represented in black. Four states of the CTD are represented: extended, loose or compacted spiral, and as a random coil (adapted from Meinhart et al. 2005).

Table 1.3 CTD interacting proteins and their functions.

Classification of the known CTD interacting proteins classified in “writers”, “readers”, and “erasers”, according to their function: CTD modification, recognising the modified CTD, or remove modification patterns. Kinases are the writers. Readers are classified into GTFs, histone and chromatin modifiers, RNA processing and with functions directed to CTD. Two heptad repeats are shown as is the sequence encompassing the functional unit. (*) Rpap2 has a dual function as is able to dock on Pol II CTD and modify Ser5-P. (Taken from Eick & Geyer 2013).

Figure 1.6 Representative distribution of CTD modification based on ChIP-seq experiments in S. cerevisea.

The signal shows the theoretical distribution of changes of CTD modification along the transcriptional cycle. Similar results have been shown for mammalian Pol II although differences have been also reported. The signals are not normalised for total Pol II abundance across the transcriptional unit. The dashed line shows the distribution of Threonine-4 in mammal cells. This figure was taken from Eick & Geyer 2013.

HAT complexes of the GNAT family														
SAGA	**SLIK**	**ADA**	**HAT-A2**	**SAGA**	**ATAC**	**PCAF**	**STAGA**	**TFTC**	**HATB**	**Elongator**	**Hpa2**
(Sc)	(Sc)	(Sc)	(Sc)	(Dm)	(Dm)	(Hs)	(Hs)	(Hs)	(Sc)	(Sc)	(Sc)
Catalytic subunit											
Gcn5	Gcn5	Gcn5	Gcn5	GCN5	GCN5	PCAF	GCN5L	GCN5L	Hat1	Elp3	Hpa2
Histones modified											
H2B/ H3/H4	H2B/ H3/H4	H3	H3	H3	H3/H4	H3/H4	H3/H4	H3/H4	H2A/H4	H3	H3/H4
Associated complex subunits											
Tra1	Tra1			TRA1		PAF400	TRRAP	TRRAP	Hat2	Elp1	Hpa2
Spt7	Spt7‡			SPT7			STAF65γ		Hif1	Elp2	
Spt8										Elp4	
Spt3	Spt3			SPT3		SPT3	SPT3	SPT3		Elp5	
Spt20	Spt20									Elp6	
Ada1	Ada1			ADA1			STAF42				
Ada2	Ada2	Ada2	Ada2	ADA2B	ADA2A	ADA2					
Ada3	Ada3	Ada3	Ada3	ADA3	ADA3	ADA3	STAF54	ADA3			
Sgf29	Sgf29	Sgf29	Sgf29	SGF29							
Sgf73	Sgf73						SCA7	SCA7			
Ubp8	Ubp8					TAF5L	TAF5L	TAF5L			
Sgf11	Sgf11					TAF6L	TAF6L	TAF6L			
Taf5	Taf5			TAF5		TAF9	TAF9	TAF9			
Taf6	Taf6			TAF6		TAF10	TAF10	TAF10			
Taf9	Taf9			TAF9		TAF12	TAF12	TAF12			
Taf10	Taf10			TAF10B				TAF2			
Taf12	Taf12			TAF12			STAF36	TAF4			
	Rtg2						STAF46	TAF5			
Chd1	Chd1							TAF6			
		Ahc1		WDA	ATAC1						
		Ahc2			HCF1						

HAT complexes of the MYST family										
NuA4	**Pic. NuA4**	**NuA3**	**SAS**	**TIP60**	**HBO1**	**MOZ/MORF**	**MSL**
(Sc)	(Sc)	(Sc)	(Sc)	(Dm/Hs)	(Hs)	(Hs)	(Dm)
Catalytic subunit							
Esa1	Esa1	Sas3	Sas2	TIP60	HBO1	MOZ/ MORF	MOF
Histones modified							
H2A/H4	H2A/H4	H3	H4	H2A/H4	H3/H4	H3	H4
Associated complex subunits							
Tra1		Yng1	Sas4	TRRAP	ING5	ING5	MSL1
Yng2	Yng2	Taf14	Sas5	ING3	ING4	BRPF1	MSL2
Yaf9		Nto1		p400	JADE1		MSL3
Eaf1				BRD8			MLE
Eaf2				EPC1			roX RNA
Eaf3				EPC2			
Eaf5				DMAP1			
Eaf6				RUVBL1	EAF6	EAF6	
Eaf7				MRG15			
Epl1	Epl1			BAF53a			
Act1				Actin			
Arp4				GAS41			
				MRGX			
				MRGBP			
				FLJ11730			
				YL1			
				TIP49a			
				TIP49b			
				TRCp120			

Table 1.4 Classification of HATs and their substrates.

The table was adapted from K. K. Lee & Workman 2007.

Histone	Phosphorylated residue	Kinases	Role
H2A	S1	?	Mitosis
	S16	rSK2	eGF signalling
	S122*(Sc)/ T120(Hs)	Bub1, NHK-1 (Dm)	DNA repair/mitosis/meiosis
	S129*(Sc)/S139(Hs,H2AX)	Mec1, Tel1 (Sc)/ATM, ATr, DNA-PK (Hs)	DNA repair
	Y142 (H2AX)	Mst1 wSTF	Apoptosis DNA repair
H2B	S10 (Sc)/S14 (Hs)	Ste20 (Sc)/Mst1 (Hs) (ipl1?)	Apoptosis Meiosis
	S32	rSK2	eGF signalling
	S36	AMPK	Transcription
	Y40 (Sc)/Y37 (Hs)	Swe1 (Sc) / wee1 (Hs)	Transcription
H3	T3	Haspin	Mitosis
	T6	PKCβ	Transcription
	S10	ipl1 (Sc)/AuroraB (Hs), rSK2, MSK1, erK1, p38, Fyn, Chk1, PrK1	Transcription, chromatin condensation, UvB response
	T11	Mek1 (Sc)/Dlk (Hs, ?) PrK1, PKM2 Chk1	Meiosis (Sc), Mitosis (Hs) Transcription DNA damage response
	S28	AuroraB, erK1/2, p38 MLTK-α, JNK1/2, MSK1	Meiosis Mitosis, transcription
	Y41	JAK2	Transcription
	T45	PK-Cδ	Apoptosis
H4	S1	CKii Sps1	DNA repair, transcription Meiosis, transcription
	S47	PAK2	(H3.3-H4) deposition
H1	S/T	CDK2	Mitosis Transcription

Table 1.5 Known kinases phosphorylating histones and their functions.
The table was adapted from Rossetto et al. 2012.

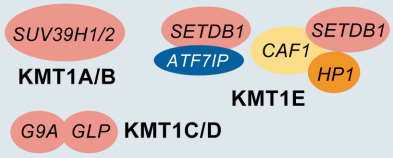
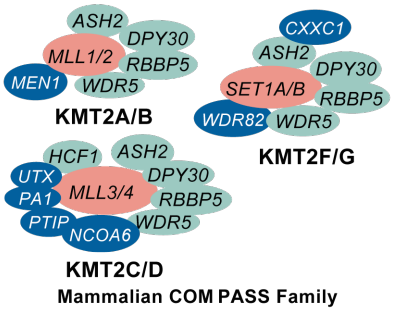
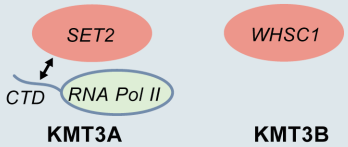
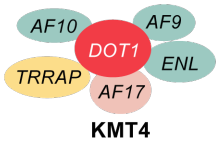
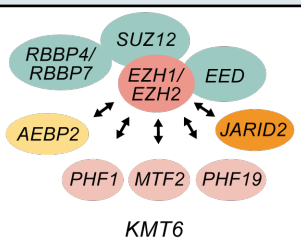
Methylase	Substrate	Mammals	*Biochemical/ biological properties*
KMT1	H3K9		*Heterochromatin formation/silencing*
KMT2	H3K4		*Active transcription; homeotic gene expression; nuclear hormone receptor signalling*
KMT3	H3K36		*Active transcription*
KMT4	H3K79		*Active transcription; cell-cycle regulation; Wnt signalling*
KMT5	H4K20		*Transcriptional repression; DNA damage response*
KMT6	H3K27		*Polycomb silencing; X chromosome inactivation; cell fate determination*

Table 1.6 Histone Lysine Methylase Complexes.

The table presented was adapted from Mohan et al. 2012.

INO80			
Subfamily	INO80	SWR1	
Complex	INO80	SRCAP	TRRAP/Tip60
Homologous subunits	hIno80	SRCAP	P400
	Tip49a, Tip49b	Tip49a, Tip49b	Tip49a, Tip49b
	BAF53a, Arp5,8	BAF53a, Arp6	BAF53a, Actin
		GAS41	GAS41
	hIes2		
	hIes6		
		DMAP1	DMAP1
		YL-1	YL-1
			Brd8/TRCp120
		H2AZ, H2B	
		ZnF-HIT1	
			TRRAP
			Tip60
			MRG15, MRGX
			FLJ11730
			MRGBP
			EPC1, EPC-like
			ING3
Unique subunits	Amida, NFRKB, FLJ20309, MCRS1, FLJ90652		

• INO80 regulates transcription and is involved in DNA repair and cell-cycle checkpoint. It has ATP-dependent nucleosome mobilization activity in vitro.
• SWR1 regulates transcription and is also involved in deposition of histone H2A.Z, chromosome stability, and cell-cycle checkpoint adaptation. In vitro it can replace pre-existing nucleosomal H2A-H2B dimers with H2A.Z-H2B dimers using ATP.

SWI/SNF	
Subfamily	SWI/SNF
Complex	BAF[f]
Homologous subunits	BRG1 or hBRM
	BAF250/ hOSA1
	BAF155, BAF170
	BAF60a
	BAF53
	hSNF5
	BAF57
	β-actin

Human BAF regulates gene expression, cell-cycle progression, organ development, and immune responses. In vitro, human BAF disrupts tailless nucleosomes. BAP and BAF contain conventional actin, similar to INO80 and SWR1

ISWI		
Subfamily	ACF/CHRAC	NURF
Complex	ACF[d]	NURF
Homologous subunits	hSNF2H	hSNF2L
	WCRF180/hACF1	BPTF
		RbAP46, RbAP48

• Subfamily ACF/CHRAC is involved in transcription, DNA replication through heterochromatin, and proper chromatin assembly. *In vitro* ACF interacts with naked DNA and nucleosomal arrays independent of ATP, assembling nucleosomes into regularly spaced chromatin.
• NURF regulates expression of homeotic genes, modulates Wnt-singaling, and affects higher-order chromatin structure; in vitro NURF catalyzes formation of regularly spaced nucleosomal arrays and facilitates transcription activation.

Mi-2/CHD		
Subfamily	CHD1/2	Mi-2/CHD
Complex	CHD1/2	NuRD[g]
Homologous subunits	CHD1 CHD2	Mi-2α/CHD3, Mi-2β/CHD4
		MBD3
		MTA1,2,3
		HDAC1,2
		RbAp46,48
		p66α,β
Unique subunits		DOC-1?

• CHD1 promotes transcription elongation; in vitro it promotes the formation of regularly spaced nucleosomal arrays.
• Mi-2/CHD deacetylates chromatin, represses transcription, and regulates development; in vitro Mi-2/CHD also promotes nucleosome sliding.

Table 1.7 Chromatin Remodelling Complexes.

The table presented was adapted from Bao & X. Shen 2007.

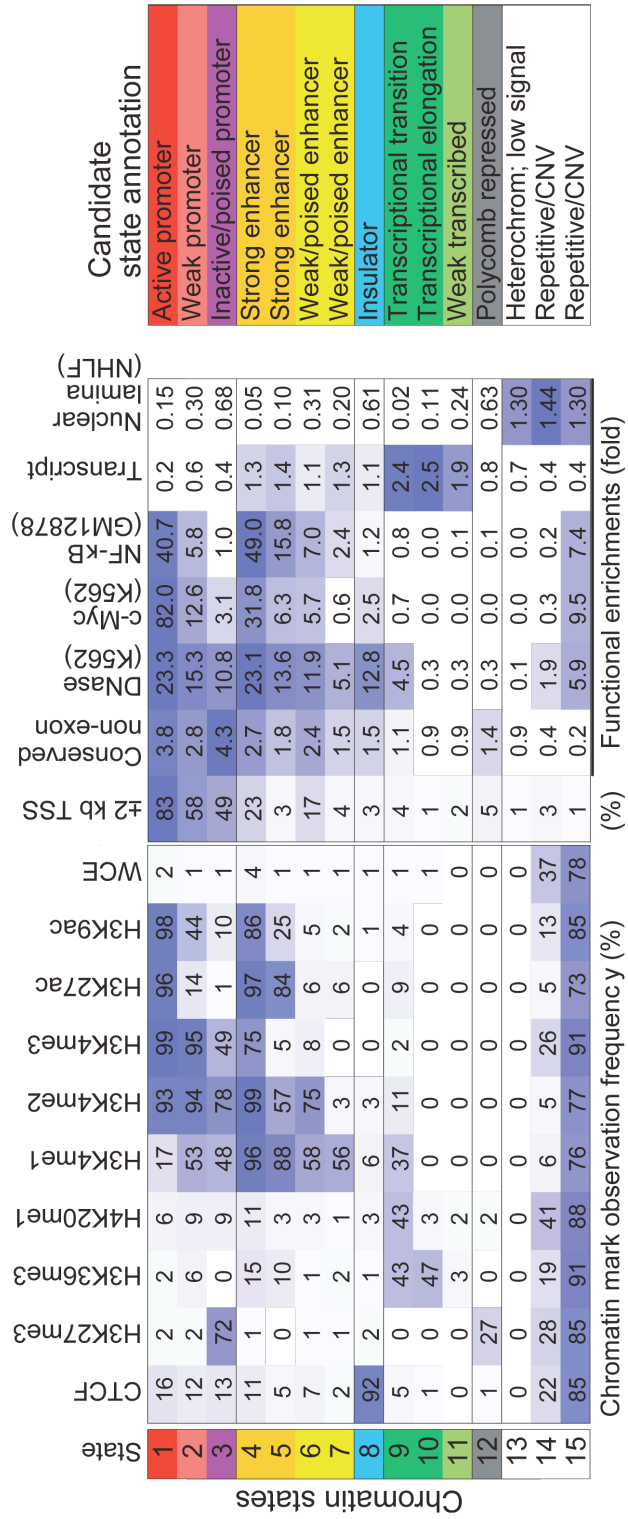


Figure 1.7 Chromatin states.

Tabular representation of the data presented in J. Ernst et al. 2011. A multivariate hidden Markov model was used to classify chromatin states across cell types. Each value represents the frequency at which a given mark is found at genomic positions corresponding to the chromatin state. The blue shading indicates the intensity, scaled by each column. WCE, whole cell extract. CNV, copy number variation (figure adapted from J. Ernst et al. 2011).

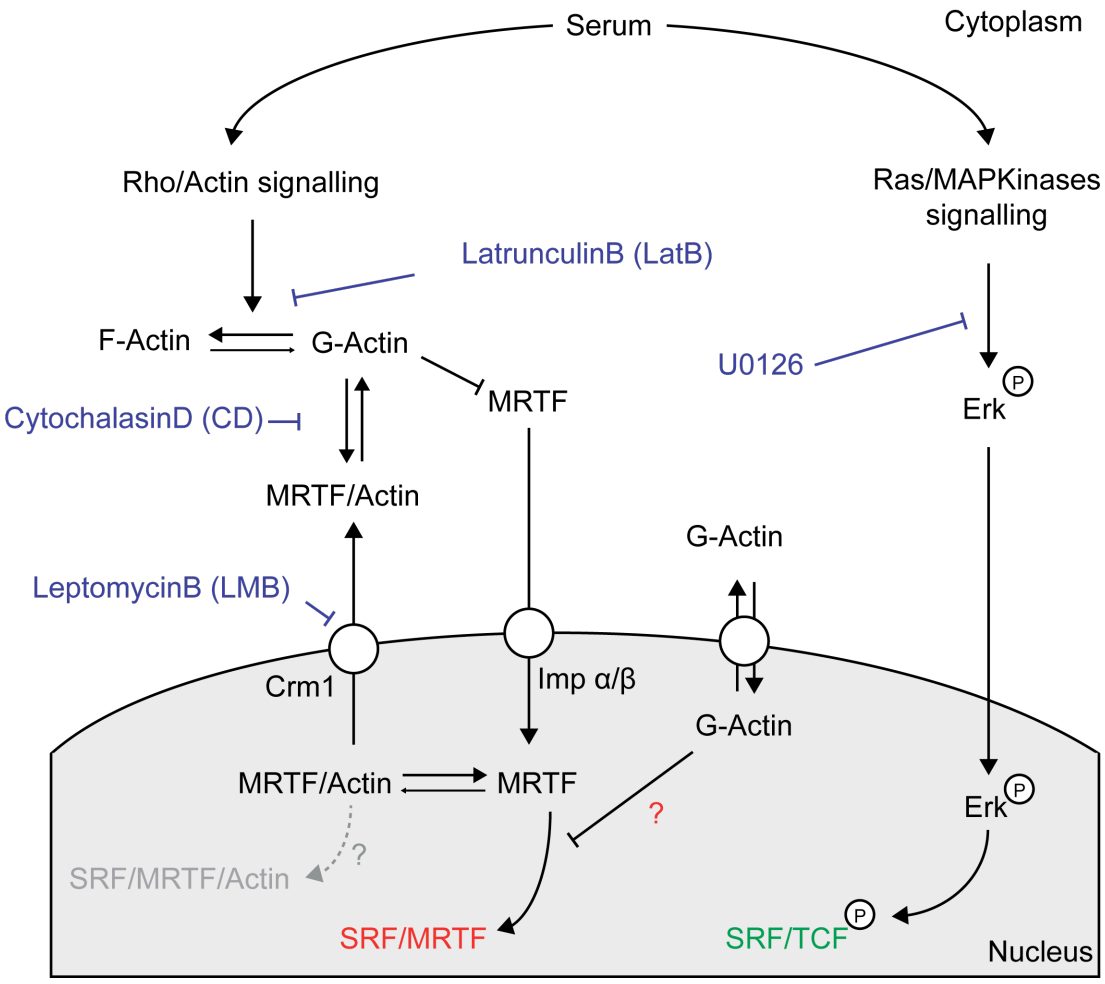


Figure 1.8 The SRF transcription network.

Serum Response Factor (SRF) controls the immediate-early transcriptional response to growth factors, and regulates numerous developmental processes including gastrulation, T-cell differentiation and muscle differentiation. To control transcription, SRF recruits signal-regulated co-activators, the Ternary Complex Factors (TCFs) and the Myocardin-Related Transcription Factors (MRTFs), which compete for a common site on its DNA-binding domain. The TCFs - SAP-1, Elk-1 and Net - are ETS proteins that link SRF activity to Ras-Erk signalling. In contrast, the two MRTFs, MRTF-A and MRTF-B, link SRF activity to Rho-actin signalling. MRTF RPEL domain acts as a G-actin sensor, controlling MRTF nuclear accumulation in response to signal-induced depletion of the G-actin pool. Agents affecting the response to either MRTFs or TCFs are shown in blue: Latrunculin B (LatB) sequesters actin monomers increasing the overall amount of monomeric actin in the cell; cytochalasin D (CD) disassociate MRTF from actin by competing with MRTF for the same binding site on monomeric actin; Leptomycin B (LMB) is a potent Crm1 inhibitor that allows MRTF nuclear accumulation in absence of signal activation; U0126 is a MEK1 and MEK2 inhibitor.

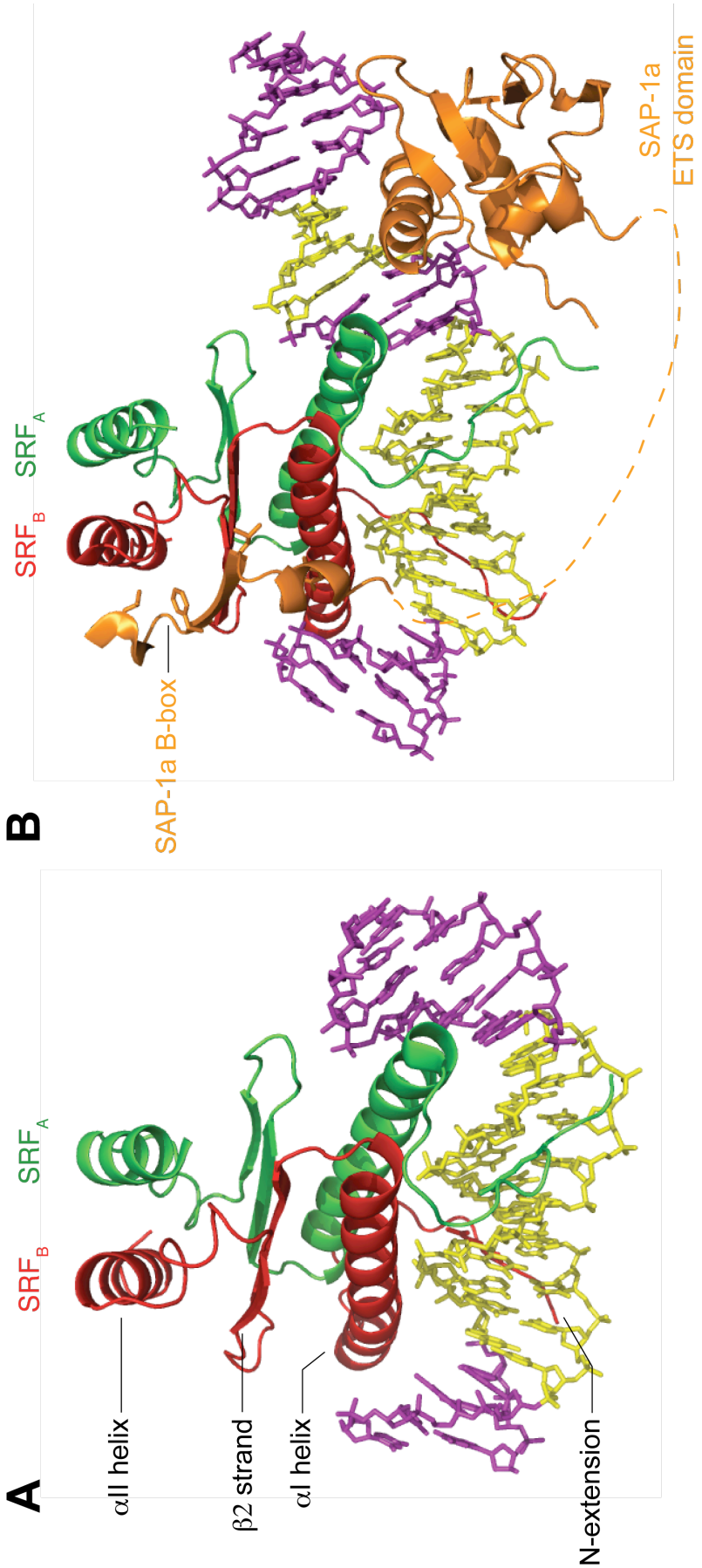

B
SRF$_B$
SRF$_A$
SAP-1a B-box
SAP-1a
ETS domain
A
SRF$_B$
SRF$_A$
αII helix
β2 strand
αI helix
N-extension

Figure 1.9 SRF structures.

The Figure was provided with permission by Sebastian Guttler. **(A)** Structure representation of the SRF DNA binding domain in combination with the SRE DNA binding consensus. The two dimers of SRF are shown in red and green in a ribbon diagram. The DNA is represented in purple with the SRE in yellow. The schematic representation was adapted from Pellegrini et al. 1995. **(B)** Representation of the structure of SRF in combination with SAP-1a. The orientation of the complex is as in panel A. SAP-1a is represented in orange. The parts of SAP-1a not resolved in the structure are represented in dashed lines and represent the linker between the B-box and the ETS domains. The ETS motif, flanking the SRE, is shown in yellow.

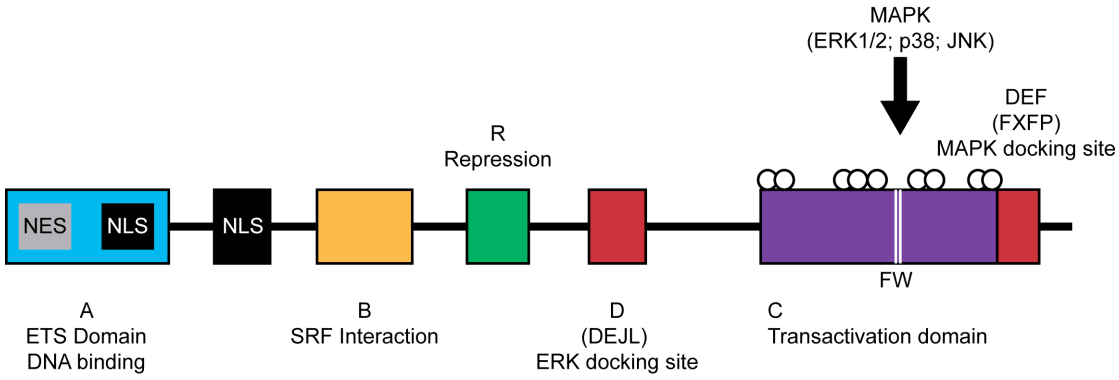


Figure 1.10 Domain organisation of proteins of the TCFs family.

The TCF subfamily of transcription factors belongs to the ETS-domain proteins and includes Elk-1, SAP-1a and Net. The domain organisation is described schematically. The ETS DNA binding domain (A, showed in light cyan) is at the N-terminal end of each TCF protein and controls the DNA binding of the TCFs. NES and putative NLS were mapped within and in close proximity of the ETS motif. The B domain (yellow) is involved in the interaction between the TCFs and SRF. The R domain (green) was reported to be involved in repression of Elk-1 transcription activity (Salinas et al. 2004). The Transactivation domain (C, showed in purple) harbours the amino acids phosphorylated by MAP kinases. Two domains surround the transactivation domain and are both responsible for the interaction with activated MAP kinases (red). The D or DEJL domain was shown to interact with active kinases of the Erk, JNK and p38 subtypes. The DEF or FXFP domain is only interacting with activated Erk. This figure was adapted from Besnard et al. 2011.

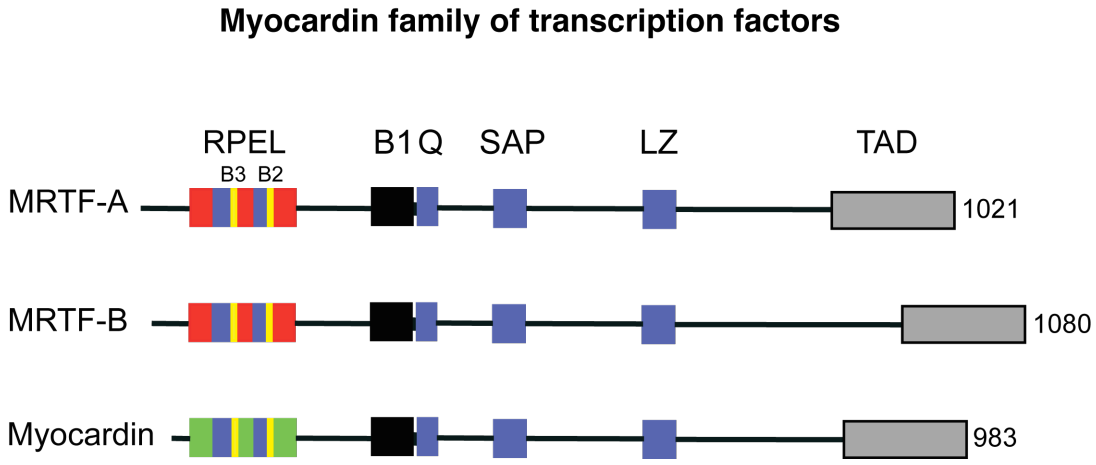


Figure 1.11 Domain organisation of proteins of the myocardin family.

The myocardin family of transcription factors is composed of myocardin (MC) and the myocardin-related transcription factors (MRTF-A and MRTF-B). RPEL motif composing the actin binding module are shown in red; B3, B2 and B1 are basic regions involved in importin interaction (B3 and B2) and in ternary complex formation (B1 region); Q, Q-rich region; SAP, SAP (SAF-A/B, acinus, pias) domain; LZ, leucine zipper required for MRTF dimerization (homo and hetero); TAD, transactivation domain heavily phosphorylated once MRTF is in the nucleus and detached from actin. Myocardin RPELs are shown in green while MRTFs' are shown in red.

Chapter 2. Definition of the transcriptional signature lead by MRTF

2.1 Aim

MRTFs are well characterised actin-binding proteins involved in the regulation of a defined set of SRF-dependent genes (Miralles et al. 2003). While the regulation of MRTF shuttling from the cytoplasm to the nucleus is relatively well understood, how they are regulated once in the nucleus is still uncharacterised. Previous work in our laboratory has shown that MRTF nuclear accumulation via Leptomycin B treatment (LMB) is sufficient for their recruitment to target promoters but disassociation from G-actin is indispensible for productive transcription (Vartiainen et al. 2007). My initial aim was to extend this analysis using genomic approaches in order to address generality and specificity. Within this chapter I am going to describe the transcriptional outcomes in response to active or inactive nuclear MRTF. I am going to characterise a specific set of genes that could be directly associated with MRTF either by function or proximity to SRF binding. I am going to describe the relationship between proximity to SRF and the level of transcription induction in response to the MRTF-activator Cytochalasin D (CD). Finally I am going to highlight the transcriptional defect observed in LMB treated cells.

2.2 Nuclear MRTF is not sufficient for both pre-mRNA and mRNA transcription

Impairment of MRTF-actin interaction by several means allows MRTF nuclear accumulation and productive activation of target genes (Miralles et al. 2003; Sotiropoulos et al. 1999). Depletion of G-actin via Rho activation or direct disruption of the MRTF-actin interaction via the MRTF-activator Cytochalasin D (CD) induces SRF-MRTF specific genes. This evidence was recently confirmed using genomic approaches (Esnault et al. 2014). The established role of actin in MRTF import and export envisioned a model where in resting condition MRTF is constantly excluded from the nucleus through Crm1-dependent export (Vartiainen et al. 2007). MRTF nuclear accumulation can be also achieved through the LMB-

dependent Crm1 blockage, fusing MRTF to Nuclear Localisations Signals (NLS) or by specifically increasing nuclear actin via NLS sequences (Posern et al. 2002; Miralles et al. 2003; Vartiainen et al. 2007). In these conditions known MRTF target genes are silenced. Any MRTF mutants unable to interact with monomeric actin will be instead competent for activation of target genes.

In order to generalise these observations I optimised a defined set of conditions, assessing for MRTF nuclear localisation and expression of a few *bona fide* targets. As showed in Figure 2.1A NIH 3T3 treated for 30 minutes with either CD or LMB results in MRTF nuclear accumulation as previously reported (Miralles et al. 2003; Vartiainen et al. 2007). As LMB is a relatively hydrophobic compound, with low solubility in aqueous solutions and a tendency to stick to plastic (see lclabs.com web site), I tested the homogeneity of treated samples. LMB was handled in glass vials and added to cell layers at 50 nM allowing for complete MRTF nuclear accumulation comparable to CD treated cells (Figure 2.1A). Target activation was followed by RNA isolation from serum-starved cells or at 30, 45 and 90 minutes after stimulation. Eight independent SRF-MRTF targets showed an acute synthesis of their first intron within 30 minutes after CD stimulation (Figure 2.1B). On the other hand LMB treated cells showed no efficient induction despite complete nuclear accumulation of MRTF. A similar transcriptional defect was also observed measuring accumulation of *Acta2* and *Srf* mature RNA (Figure 2.1C).

Both CD and LMB are pleiotropic compounds, potentially affecting other factors beside MRTF proteins. In order to nail these behaviours down to the SRF-MRTF signalling network two further conditions were analysed. 30 minutes LMB treated cells were further stimulated with either CD (LMB▶CD) or the MRTF-inhibitor Latrunculin B (LMB▶LatB). In both conditions MRTF is retained in the nucleus (Figure 2.1A). As previously reported the addition of CD allows re-activation of target genes at the level of pre- or mature RNA (Figure 2.1B and C). On the other hand the addition of LatB to LMB-treated cells does not affect either MRTF nuclear localisation or target gene activation (Figure 2.1A and C).

These conditions are going to be analysed in more details within this chapter. In conclusion, as previously reported the LMB-dependent MRTF nuclear accumulation is not sufficient for *bona fide* target gene activation. Both pre-mRNA and spliced messenger are transcriptionally defective. Furthermore LMB treated

cells could be efficiently re-induced with the actin-depolymerising agent CD, but not with LatB, which excludes unspecific effects caused by Crm1 blockage.

2.3 Genome-wide dissection of the CD and LMB response

In order to generalise the observations collected using *bona fide* SRF-MRTF targets I used RNA sequencing (RNA-seq) aiming to see whether genes induced by CD and not blocked by LMB (where LMB▶CD would show transcription activation) are transcriptionally silenced in LMB.

With the help of the bioinformatics facility I analysed both total RNA-seq reads and intronic RNA-seq reads for better sensitivity, maximising change detectability. The collected data was examined through a differential gene expression (DGE) analysis comparing each stimulus with the 0.3% FCS condition. We used the Deseq statistical method based on the negative binomial distribution (Anders & Huber 2010) in order to select genes changing in both CD and LMB▶CD conditions at an overall combined p-value less than 0.04 (each single condition scoring at p-value<0.2). 441 genes were identified showing an acute stimulation in both CD and LMB▶CD conditions after stimulation (Figure 2.2A and B). As shown in Figure 2.2C LMB treatment does not activate efficient transcription of these targets and does not affect their re-activation as shown by the LMB▶CD condition. Furthermore the addition of LatB to LMB showed no change if compared to the LMB condition. Such an unbiased approach allowed me to specifically select genes activated by CD regardless of any possible LMB-unspecific effect.

2.4 CD-activated genes are functionally related to MRTF

I next tested if the selected 441 genes could be directly linked to SRF-MRTF activities. As previously reported in the laboratory SRF-MRTF targets are significantly enriched in genes involved in actin filament dynamics, cell adhesion, cell motility, and other actin-linked processes (Esnault et al. 2014). I therefore performed a gene ontology (GO) analysis of the 441 selected targets using DAVID in order to assess any enrichment of similar functional classes (Huang et al. 2009). It was possible to reveal several genes involved in actin filament processes including focal adhesion, stress fibers, actomyosin complexes and contractile parts (Figure 2.3A). Furthermore a group of 48 genes defined a class of targets involved

in regulation of transcription including several transcription factors and co-regulators controlling cell differentiation, morphogenesis and motility.

To further examine the specificity of the described signature I directly compared the functional terms enriched in this analysis with the SRF-MRTF signature previously described by us (Esnault et al. 2014). As shown in Figure 2.3B there is a significant overlap in functional signatures associated with SRF and MRTF specific genes. From this first analysis it is possible to conclude that genes specifically activated by CD and unaffected by LMB are functionally related to SRF and MRTF target genes previously classified in the laboratory.

2.5 CD-activated genes are associated with SRF binding

To further characterise the association between SRF-MRTF to the selected 441 target genes I performed an unbiased analysis of the distance to SRF of each selected target using the SRF binding distribution described previously in the lab (Esnault et al. 2014). As it was previously reported SRF binding occurs preferentially within 70 Kb of actively transcribed genes (Esnault et al. 2014). To assess if such a relationship was also observed within the analysed dataset, I examined the distance of active and inactive genes from the SRF binding sites defined in our previous work (Esnault et al. 2014). I defined active genes as those showing at least five RNA-seq reads mapping within intronic features and unaffected by the stimuli used. SRF-binding sites were again found most frequently within 70Kb of active genes, confirming our previous observation (Figure 2.4A and D).

I then proceeded to analyse the average distance to SRF of the 441 targets compared to active un-induced genes or to genes that were scoring as induced only in one out of two conditions at p-value less than 0.2 (CD only and LMB▶CD only) (Figure 2.4B). As shown, only the 441 targets are overall significantly closer to SRF if compared to active genes. Furthermore genes induced only by CD or LMB▶CD are not closer to SRF than active un-induced genes. As it is possible to see in Figure 2.4C more than 60% of the 441 genes have an SRF binding within 70 Kb from their TSS and 53% of these are proximal to an SRF binding event (35% of the 441 are within 2 Kb of an SRF binding site). In addition the distribution between “direct” (within 2Kb of an SRF site) and “near” (within 70Kb of an SRF site) for

genes responding only to CD or LMB▶CD is not significantly different from active un-induced genes (Figure 2.4C).

In conclusion genes induced by CD and unaffected by LMB, besides being functionally related to SRF-MRTF target genes, are physically linked to SRF.

2.6 Nuclear accumulation of MRTF via LMB is transcriptionally defective

To further characterise whether it was possible to see any LMB-dependent induction of MRTF-specific genes I further assessed whether any of the 441 genes were statistically upregulated following LMB treatment. A group of 199 genes, overall closer to SRF, were significantly upregulated if compared to resting conditions (Figure 2.5A). 141 out of 199 were within 70kb of an SRF binding but the induction observed was still significantly reduced if compared to CD treated cells (Figure 2.5 B). Furthermore this induction showed LatB sensitivity only at the level of intronic-reads (Figure 2.4B). This observation is consistent with unpublished data collected in the lab showing a mild enhancement of a few highly expressed SRF-MRTF targets following LMB stimulation (unpublished data discussed in Sebastian Guettler thesis 2007).

With the help of the bioinformatics facility I assembled density profiles of RNA-seq data averaging reads mapped in intronic features - using the 288 CD induced genes within 70kb of an SRF binding site (Figure 2.5C). In CD induced conditions it is possible to see a strong accumulation of reads towards the 5' end with a progressive reduction towards the 3' end of the genes (Figure 2.5C). Such a profile could only partially provide information about travelling Pol II and in this context I used these profiles to measure the transcriptional induction relative to the distance from the TSS. It is possible to confirm that LMB treatment does not induce efficient transcription. Disassociation of MRTF from nuclear actin via CD is required to induce synthesis of precursor RNA.

I further characterised the MRTF-dependent gene induction by assessing the RNA fold induction in CD over 0.3% FCS and its relation with the distance to SRF sites. As shown in Figure 2.6A highly induced genes are closer to SRF binding sites (Figure 2.6A). Such a relationship was only observed using intronic-reads as opposed to total-reads, suggesting a direct relation between stimulation

and RNA precursor synthesis (Figure 2.6B). Furthermore it was possible to observe a partial induction of intronic-reads in LMB of highly induced genes (Figure 2.6A). This induction showed LatB sensitivity only for the top 40% induced genes (114 genes). The addition of CD following LMB stimulation showed enhanced production of intronic-reads, rescuing the observed defect (Figure 2.6A). Despite the clear induction, the LMB▶CD condition showed reduced expression if directly compared with the CD condition.

In order to see whether the relation between inducibility and SRF proximity was specific for SRF-MRTF controlled genes I examined the distribution of the distances to SRF for LMB-specific induced genes (Figure 2.6C). It was possible to select 758 genes within 70Kb of an SRF binding site showing a mild induction after LMB treatment (Figure 2.6C). No relation between inducibility and SRF binding was observed for this group of targets. The overall level of induction of these genes was less significant than the one observed in CD for SRF-MRTF specific genes.

In conclusion MRTF nuclear accumulation does not promote efficient gene expression. A slight increase in intronic reads, which might reflect inefficient RNA synthesis, occurs only at highly CD-induced genes as previously reported in the lab (unpublished data). Given that such an LMB-dependent induction is sensitive to LatB treatments, it can be led back to MRTF-dependent nuclear activities.

2.7 Summary

This chapter has extended previous observations collected and published by the lab (Vartiainen et al. 2007). MRTF nuclear accumulation via LMB is not sufficient to activate productive transcription of *bona fide* SRF-MRTF targets. LMB is not unspecifically blocking transcription as re-addition of CD allows expression of selected targets. This transcriptional signature is specific for SRF-MRTF genes as targets fulfilling these criteria are functionally related to SRF-MRTF specific targets. Furthermore genes specifically activated by CD and unaffected by LMB are overall closer to previously defined SRF binding sites (Esnault et al. 2014). Induction of precursor RNA, observed in transcriptionally competent conditions, correlates with the distance to SRF. A mild induction of SRF-MRTF targets after LMB treatment is observed only at intronic-reads and is sensitive to the MRTF inhibitor LatB.

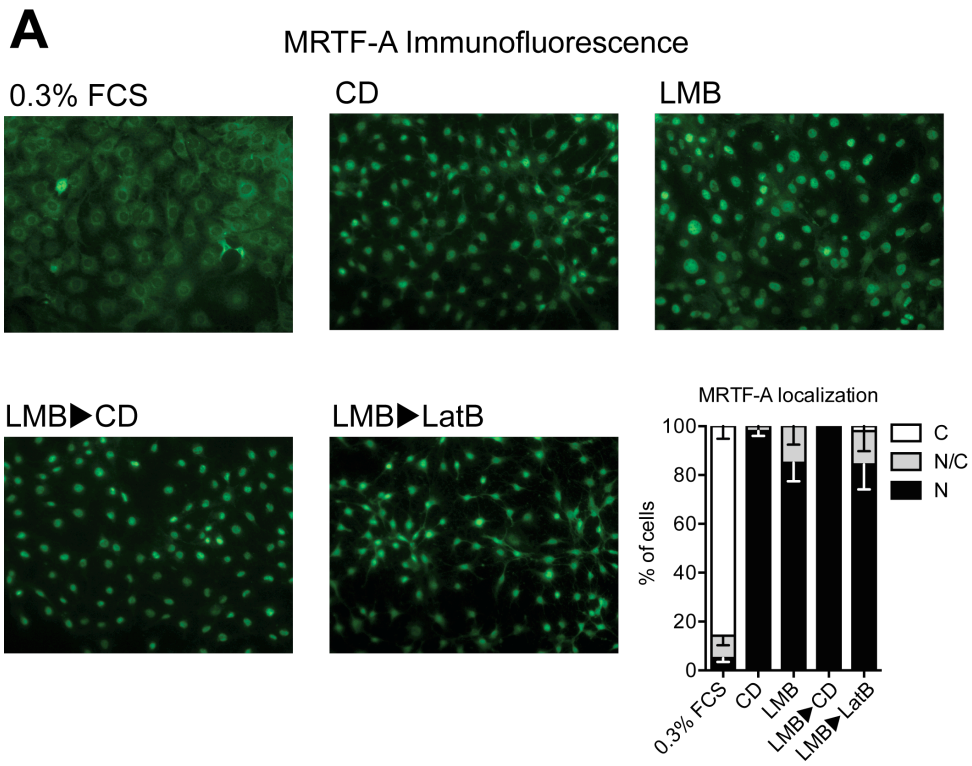

A
MRTF-A Immunofluorescence
0.3% FCS
CD
LMB
LMB▶CD
LMB▶LatB
MRTF-A localization
% of cells
C
N/C
N
0.3% FCS
CD
LMB
LMB▶CD
LMB▶LatB
B
pre-mRNA accumulation time course
Acta2
Average
(n=8)
Acta2
Average
(n=8)
RNA abundancy/Gapdh
% of the max
Time (min)
C
mRNA accumulation time course
Acta2
Srf
Acta2
Srf
RNA abundancy/Gapdh
Time (min)
CD
LMB
LMB▶CD
LMB▶LatB

Figure 2.1 MRTF nuclear localisation is not sufficient for target activation.

(A) Immunofluorescence microscopy of MRTF-A in NIH3T3 fibroblasts. Cells were serum-starved and stimulated as indicated (CD, cytochalasin D; LMB, leptomycin B; LMB▶CD, 30 minutes LMB followed by 30 minutes CD; LMB▶LatB, 30 minutes LMB followed by 30 minutes Latrunculin B). Bottom right of panel A is the quantification where 150-300 cells were scored according to predominantly nuclear (N), pancellular (N/C) and predominantly cytoplasmic (C) MRTF-A localisation. **(B)** Accumulation of pre-mRNA of *bona fide* SRF-targets. NIH3T3 fibroblasts were serum-starved overnight and then stimulated with CD (blue), LMB (purple), LMB▶CD (orange) or LMB▶LatB (red) for 30, 45 or 90 minutes. RNA was isolated and abundance of the first intron of the SRF target genes *Acta2*, *Srf*, *Cyr61*, *Vcl*, *Msrb3*, *Dstn*, *Ctgf* and *Sorb1* analysed by quantitative RT-PCR. Relative abundances obtained after normalisation for *Gapdh* message for *Acta2* and an average profile normalised to the highest signal is shown. Data represent technical triplicates with SEM. **(C)** Accumulation of mature mRNA of *Acta2* and *Srf* following the same time course as in panel B. Primers spanning an exon1-exon2 junction were used for both targets.

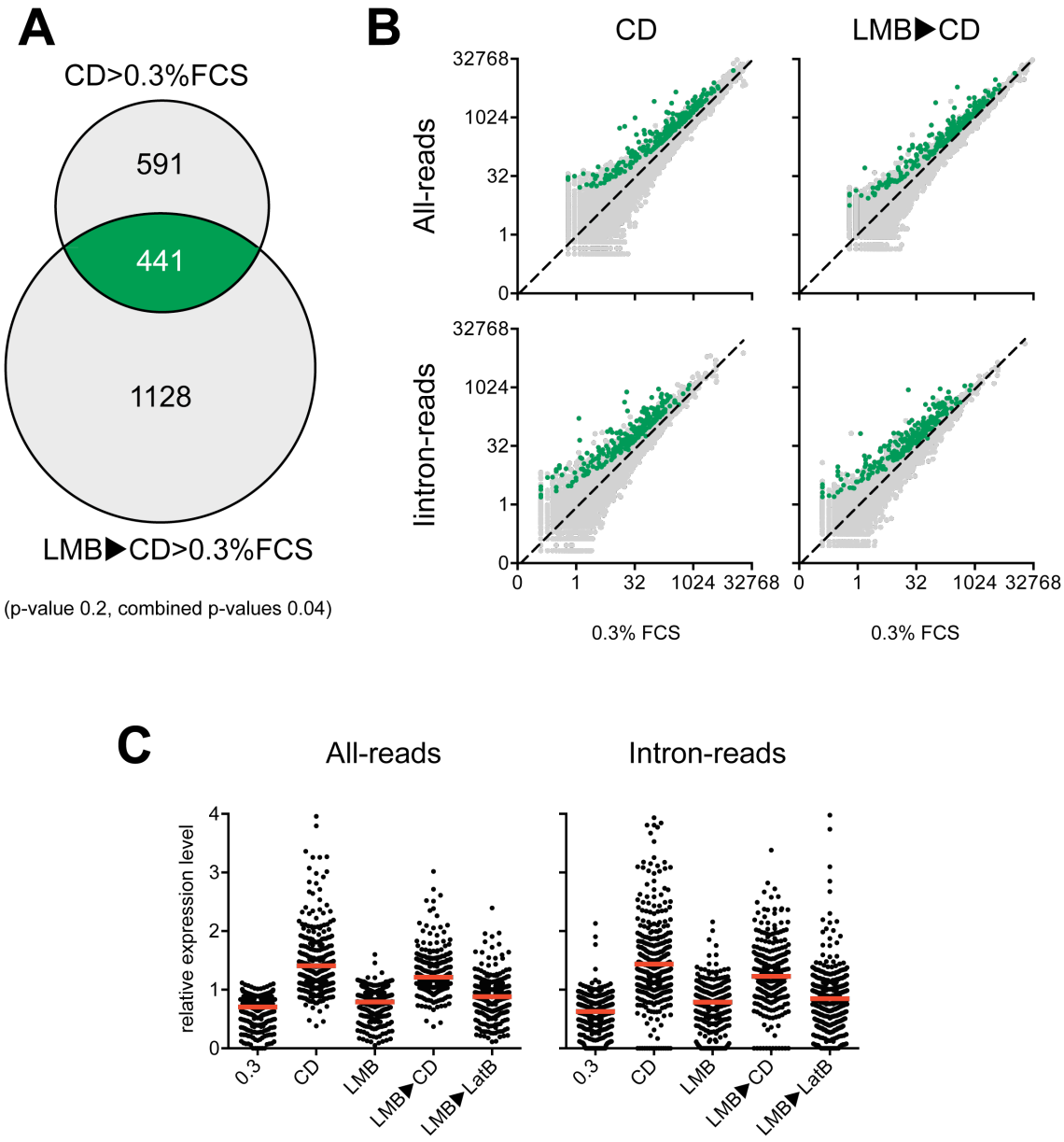


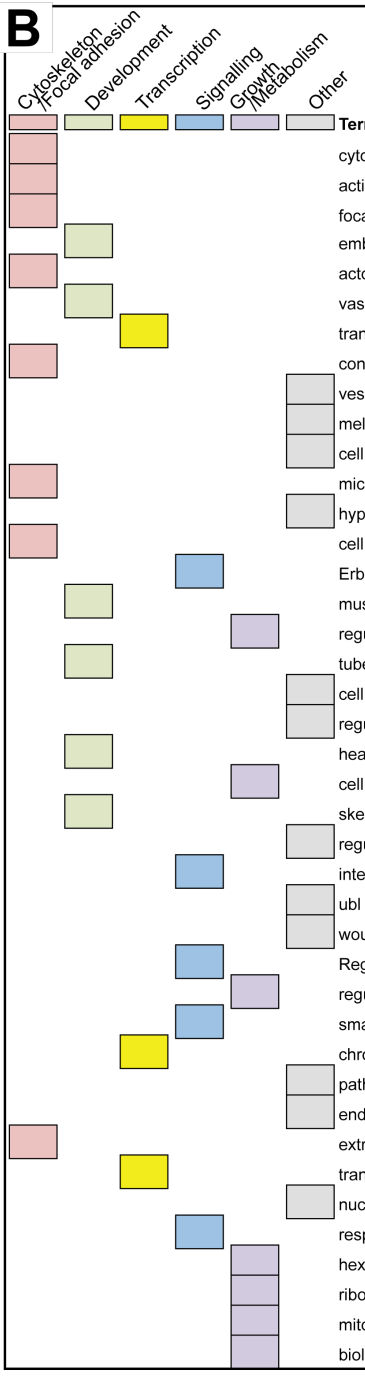
Figure 2.2 LMB stimulation is defective compared to CD stimulation

(A) Venn diagram showing genes induced at the intronic or total reads in either CD or LMB▶CD if compared to 0.3% FCS (p-value<0.2 per single comparison). The green quadrant highlights 441 genes that are significantly induced in both CD and LMB▶CD if compared to resting condition at a combined p-value<0.04. **(B)** Scatter plot display of total (*top*) and intronic (*bottom*) RNA-seq read counts before and after CD (*left*) or LMB▶CD (*right*) stimulation. The 441 genes induced in both conditions are highlighted in green. **(C)** Display of the relative induction per gene across conditions. (Red bar) Median. Total (*left*) or intronic (*right*) RNA-seq read counts were normalised to the average across conditions.

A

Term	PValue	n	FDR	Bonferroni	Benjamini
Actin filament based process	2.3E-13	26	3.9E-10	3.9E-10	3.9E-10
Basolateral plasma membrane	1.4E-09	19	1.9E-06	4.5E-07	1.1E-07
Cytoskeleton	2.7E-09	40	3.6E-06	8.7E-07	4.3E-07
Focal adhesion	3.0E-09	22	3.4E-06	3.6E-07	3.6E-07
Adherens junction	8.2E-09	16	1.1E-05	2.6E-06	3.7E-07
Contractile fiber part	2.3E-06	12	3.2E-03	7.4E-04	7.4E-05
Cell cortex	1.8E-05	13	2.5E-02	5.7E-03	4.8E-04
Regulation of cytoskeleton organization	8.6E-05	11	1.4E-01	1.4E-01	2.9E-02
Stress fiber	1.1E-04	6	1.5E-01	3.5E-02	2.1E-03
Regulation of organelle organization	2.1E-04	13	3.6E-01	3.0E-01	3.6E-02
Actomyosin	2.5E-04	6	3.4E-01	7.6E-02	3.8E-03
Transcription regulator activity	4.9E-04	48	6.9E-01	2.0E-01	5.5E-02

B



Term	Esnault et al. 2014 SRF (n=960) P-value	n	MRTF (n=921) P-value	n	CD and LMB▶CD responsive targets (n=441) P-value	n
cytoskeleton	3.E-15	115	3.E-15	112	3.E-09	40
actin cytoskeleton organization	4.E-13	39	7.E-13	38	4.E-13	25
focal adhesion	7.E-11	19	4.E-11	19	3.E-09	13
embryonic development	2.E-06	47	2.E-06	46	8.E-03	19
actomyosin	3.E-06	10	2.E-06	10	3.E-04	6
vasculature development	7.E-06	32	9.E-06	31	5.E-03	14
transcription regulator activity	2.E-05	159	9.E-06	156	5.E-04	48
contractile fiber	5.E-05	15	3.E-05	15	ns	ns
vescicle-mediated transport	5.E-05	46	2.E-05	46	ns	ns
melanosome	5.E-05	15	2.E-05	50	3.E-02	14
cell projection	6.E-05	50	4.E-05	15	2.E-06	12
microtubule cytoskeleton	1.E-04	41	1.E-04	41	3.E-05	26
hypertrophic cardiomyopathy (HMC)	1.E-04	15	7.E-05	15	ns	ns
cell junction	3.E-04	41	2.E-04	39	ns	ns
ErbB signaling pathway	5.E-04	14	6.E-04	21	7.E-03	11
muscle organ development	9.E-04	21	9.E-04	34	4.E-02	7
regulation of fat cell differentiation	1.E-03	6	9.E-04	27	ns	ns
tube development	2.E-03	27	1.E-03	6	ns	ns
cell motion	2.E-03	34	1.E-03	13	ns	ns
regulation of nucleocytoplasmic transport	2.E-03	8	3.E-03	6	ns	ns
heart development	3.E-03	23	4.E-03	14	ns	ns
cell cycle	3.E-03	49	5.E-03	22	ns	ns
skeletal system development	4.E-03	27	5.E-03	11	ns	ns
regulation of endothelial cell proliferation	4.E-03	6	5.E-03	26	ns	ns
integrin-mediated signaling pathway	6.E-03	11	7.E-03	28	ns	ns
ubl conjugation	6.E-03	41	8.E-03	7	ns	ns
wound healing	6.E-03	14	7.E-03	46	ns	ns
Regulation of MAP kinase activity	7.E-03	14	8.E-03	39	ns	ns
regulation of growth	8.E-03	24	9.E-03	56	6.E-03	31
small GTPasemediated signaltransduction	9.E-03	24	1.E-02	13	4.E-02	6
chromatin modification	1.E-02	22	1.E-02	15	ns	ns
pathways in cancer	1.E-02	28	1.E-02	23	ns	ns
endosome	1.E-02	19	2.E-02	22	ns	ns
extracellular matrix part	2.E-02	12	2.E-02	16	ns	ns
transcription factor activity	2.E-02	56	2.E-02	21	ns	ns
nuclear envelope	2.E-02	15	2.E-02	16	ns	ns
response to hormone stimulus	3.E-02	16	3.E-02	17	ns	ns
hexose metabolic process	3.E-02	16	3.E-02	11	ns	ns
ribonucleoprotein complex	ns	ns	ns	ns	ns	ns
mitochondrion	ns	ns	ns	ns	ns	ns
biological rhythms	ns	ns	ns	ns	ns	ns

Figure 2.3 Go analysis of CD-specific induced genes.

(A) DAVID analysis of genes induced by CD and LMB▶CD. Functional classes enriched with a minimum p-value of 10^{-4} are highlighted in red. Post-hoc analysis (Bonferroni and Benjamini) is shown. **(B)** Comparison of functional classes enriched within the 441 data set (CD and LMB▶CD) and data collected in Esnault et al 2014 (SRF, MRTF) (Esnault et al. 2014).

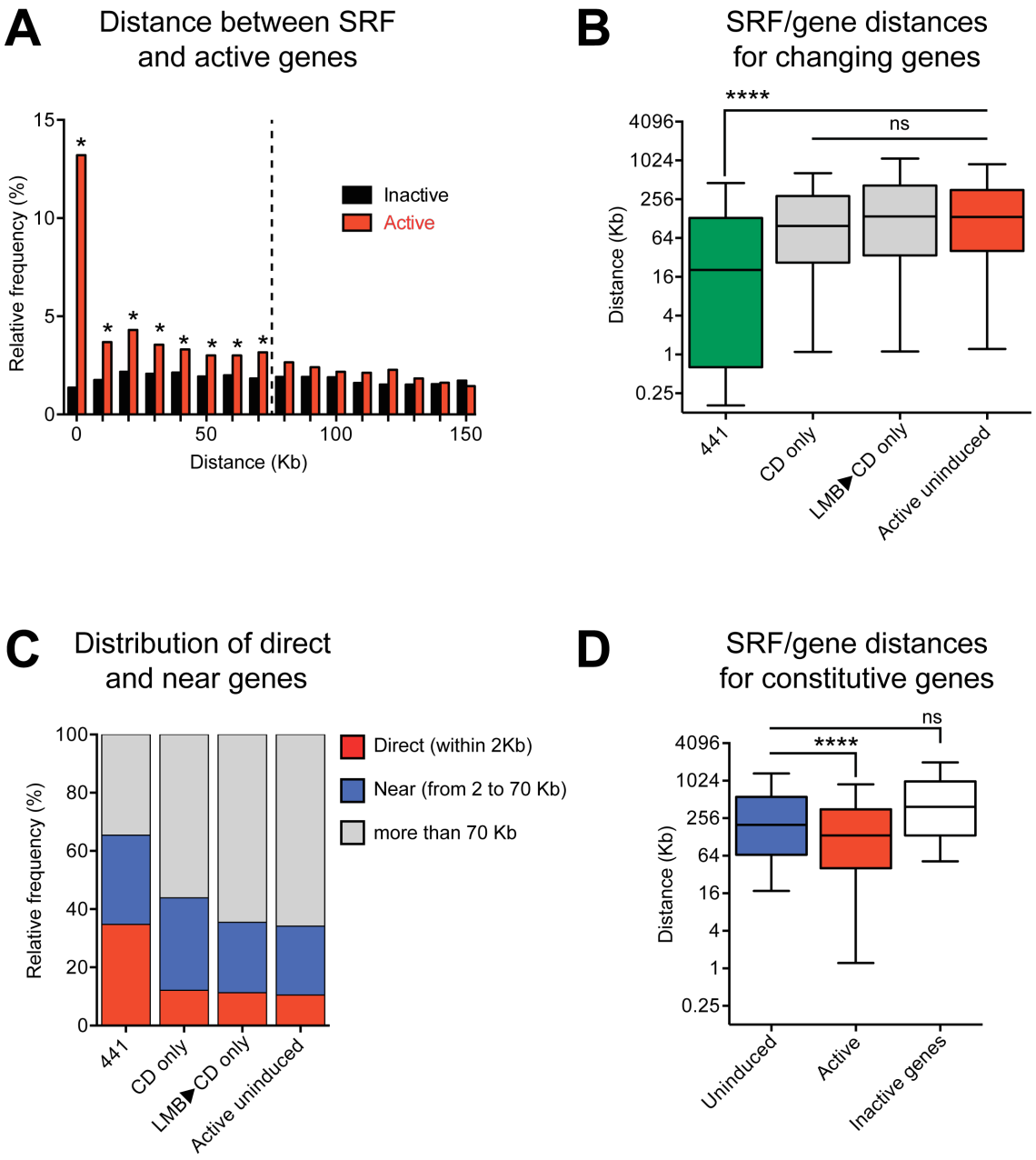

A
Distance between SRF and active genes
Relative frequency (%)
Distance (Kb)
Inactive
Active
B
SRF/gene distances for changing genes
Distance (Kb)
****
ns
441
CD only
LMB▶CD only
Active uninduced
C
Distribution of direct and near genes
Relative frequency (%)
Direct (within 2Kb)
Near (from 2 to 70 Kb)
more than 70 Kb
441
CD only
LMB▶CD only
Active uninduced
D
SRF/gene distances for constitutive genes
Distance (Kb)
ns
****
Uninduced
Active
Inactive genes

Figure 2.4 Genes induced by CD and unaffected by LMB are spatially linked to SRF binding sites.

(A) SRF sites are overrepresented within 70kb of transcriptionally active genes. The graph represents the frequencies of SRF sites relative to active and inactive genes (per 10kb bin relative to TSS) using coordinates of SRF sites published in Esnault et al. 2014. (Asterisks *) Significant at $P < 0.05$, multiple t-test with Holmes-sidak correction. **(B)** 441 induced genes are significantly closer to SRF than constitutively active and genes that score as induced in CD only or LMB▶CD only. (Asterisks****) significant at $P < 0.0001$, Mann-Whitney test; (ns) non significant **(C)** Distribution of the distances to SRF per gene group. (Direct) Sites within 2Kb of 5'-flanking sequences; (near) sites within 70kb of the TSS. **(D)** Distances distribution of SRF sites to uninduced genes. (Blue) All uninduced genes (n = 16385); (red) active un-induced genes with at least 5 RNA-seq read counts in intronic features (n = 7264); (white) inactive genes with no reads in intronic features (n= 5062). (Asterisks) significant at $P < 0.0001$, Mann-Whitney test; (ns) non significant. For the box plot graphs in panel B and D the middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90^{th} and 10^{th} percentiles.

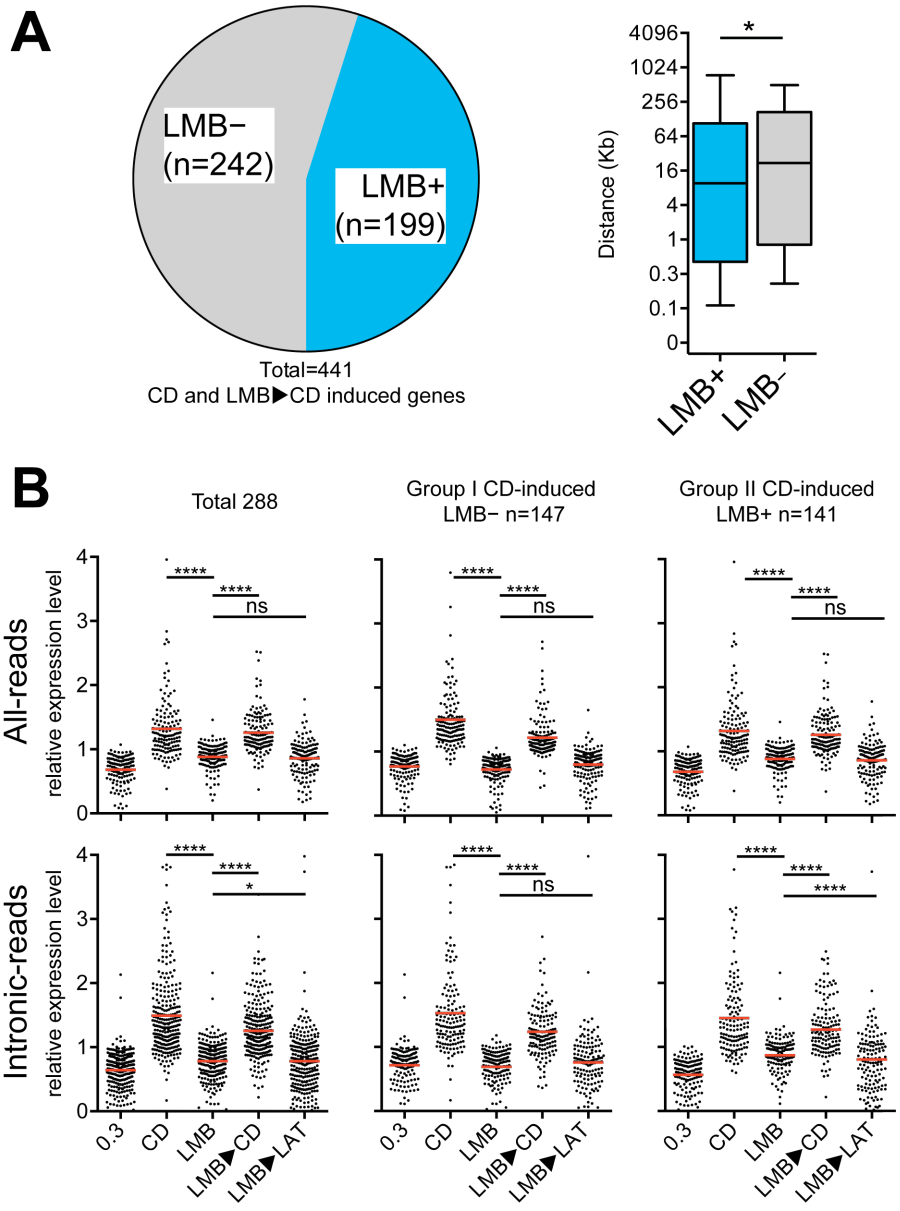

A
LMB−
(n=242)
LMB+
(n=199)
Total=441
CD and LMB▶CD induced genes
Distance (Kb)
4096
1024
256
64
16
4
1
0.3
0.1
0
*
LMB+
LMB-
B
Total 288
Group I CD-induced
LMB− n=147
Group II CD-induced
LMB+ n=141
All-reads
Intronic-reads
relative expression level
****
ns
*
0.3
CD
LMB
LMB▶CD
LMB▶LAT
C
Intronic-reads meta profiles
Normalised count per base
0.3% FCS
CD
LMB
LMB▶CD
LMB▶LatB
intron contigs

Figure 2.5 LMB stimulation does not induce efficient activation of MRTF-specific genes.

(A) (*Left*) partitioning of the 441 CD/LMB▶CD induced genes in statistically changing in LMB (LMB+) or non-changing in LMB (LMB-) (p-value < 0.2). (*Right*) SRF binding sites are closer to genes showing change also in LMB. (Asterisk*) Significant at P < 0.05, Mann-Whitney test. For the box plot graphs the middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90^{th} and 10^{th} percentiles. **(B)** Display of the relative induction per gene within 70Kb of an SRF binding site across conditions. All-reads (*top row*) or intronic (*bottom row*) RNA-seq read counts were normalised to the average across conditions. CD/LMB▶CD induced genes within 70Kb of an SRF binding site (Total 288, left column); CD/ LMB▶CD induced genes within 70kb non-changing in LMB (Group I LMB–, central column) and CD/LMB▶CD induced genes within 70kb changing in LMB (Group II LMB+, right column). (*) Significant at P < 0.05, (****) significant at P<0.0001, Wilcoxon test; (ns) non significant. **(C)** RNA-seq density profiles of intronic RNA-seq reads, for CD/LMB▶CD induced genes within 70kb of an SRF site (n = 288). 0.3% FCS (black line), CD (blue line), LMB (purple line), LMB▶CD (orange line) and LMB▶LatB (red line).

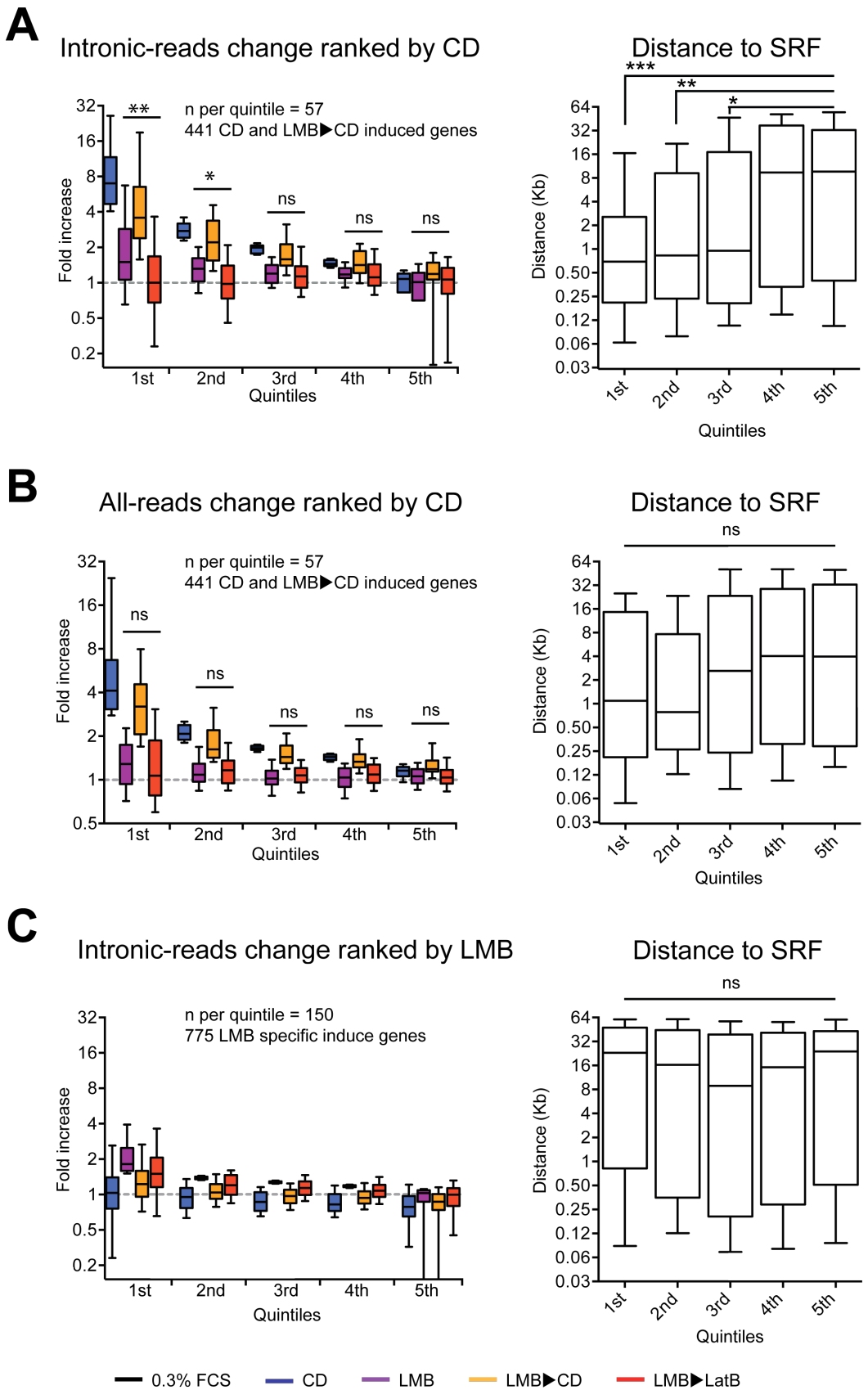

A
Intronic-reads change ranked by CD
n per quintile = 57
441 CD and LMB▶CD induced genes
Fold increase
Quintiles
1st
2nd
3rd
4th
5th
**
*
ns
Distance to SRF
Distance (Kb)
***
B
All-reads change ranked by CD
C
Intronic-reads change ranked by LMB
n per quintile = 150
775 LMB specific induce genes
0.3% FCS
CD
LMB
LMB▶CD
LMB▶LatB

Figure 2.6 Highly induced genes are closer to an SRF site and show partial LMB-dependent stimulation.

(A) Highly induced genes show partial response to LMB at the intronic RNA-seq reads level and greater proximity to SRF binding sites. (*Left*) Quintiles ranked by CD induction compared to resting 0.3% FCS condition. Each condition is plotted per quintile. Asterisks shows comparison between LMB and LMB▶LatB. (*) P < 0.05; (**) P < 0.01; (ns) non significant, Wilcoxcon test. (*Right*) Distribution of the distances to SRF binding sites per quintile. (*) P < 0.05, (**) P< 0.01, (***) P < 0.001, Mann-Whitney test. **(B)** As panel A but using total RNA-seq reads ranked by CD induction over 0.3% FCS. (ns) non significant, Wilcoxcon test and Mann-Whitney test for the right panel. **(C)** LMB induced genes are not linked to SRF. (*Left*) LMB fold induction ranking of genes within 70kb from an SRF binding site specifically changing only with LMB stimulation (n = 775). (*Right*) Distribution of the distances to SRF binding sites per quintile. (ns) non-significant, Mann-Whitney test. For the box plot graphs in each panel the middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90^{th} and 10^{th} percentiles.

Chapter 3. Nuclear actin controls MRTF DNA binding

3.1 Aim

MRTF nuclear activities are directly affected by monomeric actin. As presented in Chapter 2 nuclear accumulation of MRTF is not sufficient for productive transcription of SRF-MRTF targets genome-wide. Efficient disassociation of G-actin from MRTF is essential for accumulation of both precursor and messenger RNAs. Furthermore MRTF nuclear accumulation under resting conditions does not affect transcriptional reactivation of MRTF specific genes, allowing us to exclude any unspecific transcriptional impairment.

The mechanism by which MRTF activates transcription is still unclear. Recently we have been able to describe how serum activates the SRF signalling pathway and in particular how MRTFs are affecting Pol II recruitment and elongation (Esnault et al. 2014). A rigorous analysis of the signal-dependent Pol II recruitment allowed us to observe that MRTF DNA binding promotes both Pol II recruitment and escape according to the gene context (Esnault et al. 2014). Although required, the association of MRTF to target promoters seems not sufficient for productive transcription as shown in our laboratory for some bone fide MRTF-specific promoters (Vartiainen et al. 2007).

Within this chapter I am aiming to elucidate, using both genomic and biochemical approaches, how MRTF-DNA binding is controlled. This chapter will focus on two central questions: is MRTF nuclear accumulation sufficient for MRTF-DNA binding genome-wide, and does nuclear actin affect MRTF-DNA association?

3.2 Actin affects MRTF DNA binding and SRF-MRTF cooperative binding

Nuclear MRTF activities might be controlled at the level of DNA binding. Although preliminary observations collected in the laboratory have shown that MRTF nuclear accumulation is sufficient for its association with Srf, Vcl and Cyr61 promoters, it is not yet clear if actin affects MRTF-DNA binding. I used chromatin immunoprecipitation (ChIP) combined with deep sequencing (ChIP-seq) to define,

at the genome scale, both MRTF-A and SRF binding behaviours (Figure 3.1A and 3.2A). The conditions described in chapter 2 where used so that DNA binding and gene activity could be correlated. Instead of defining de-novo a set of SRF and MRTF binding sites I analysed my data in relation to our already defined set of SRF and MRTF-binding sites (Esnault et al. 2014). As published by us, in fibroblasts it is possible to define a core set of 3133 SRF-binding sites and 2416 MRTF-binding sites (Esnault et al. 2014). 95% of the MRTF-binding sites are associated with SRF and two thirds of the SRF binding relies on nuclear MRTF. Despite what was reported in the past (Herrera et al. 1989), the majority of SRF-binding sites show cooperative binding with nuclear MRTF. LatB-dependent inhibition of MRTF impairs both MRTF and SRF binding at inducible sites. Inducible and constitutive sites could be defined using an inducibility threshold such that the linear regression curve was closer to 1 for the constitutive sites.

With the help of the bioinformatics facility I defined peaks scoring as significantly enriched over background using a MACS threshold of $P<10^{-5}$ for both SRF and MRTF-A (Y. Zhang et al. 2008b). We defined an SRF data set based on the 3133 SRF-binding sites published by the lab (Esnault et al. 2014). 2547 SRF binding sites, detected in at least one out of five analysed conditions (0.3% FCS, CD, LMB, LMB▶CD and LMB▶LatB), besides being called by MACS, coincides with the 3133 set. Using this core of SRF peaks I defined de-novo inducible and constitutive SRF binding sites using the CD stimulation as a reference. A threshold of 1.9 fold induction allowed me to define constitutive and inducible sites with a mean inducibility of 1.01 ± 0.01 and 3.1 ± 0.05 respectively. I then asked whether I could see any inducibility in CD in the remaining 586 SRF-binding sites not called by MACS. 115 peaks showed inducible behaviour in CD if compared to 0.3% FCS. Out of the remaining 471 SRF-binding sites 104 showed no signal in any condition while the remaining 367 showed ChIP-seq read distribution with good correlation if compared with our previous dataset. I therefore included this final group obtaining a data set of 3029 SRF-binding sites. 1604 MRTF-A binding sites were defined as scoring in at least one out of the four conditions where MRTF-A is nuclear (CD, LMB, LMB▶CD and LMB▶LatB), and coinciding with the 3133 data set. In addition I considered as MRTF associated those binding sites where SRF signal was inducible in CD, obtaining a set of 2188 MRTF-A sites.

I then analysed the binding behaviour of MRTF-A across the defined 2124 binding sites in the different conditions used. MRTF-A nuclear accumulation via LMB showed to be sufficient for DNA binding at constitutive and inducible binding sites as shown for the *Acta2* and *Cofilin* promoters (Figure 3.1A). The binding intensity in LMB compared to CD was 3 to 4 times weaker (Figure 3.1B). Strikingly the addition of CD or LatB changed the MRTF-A signal dramatically across all the 2124 binding sites (Figure 3.1B). In particular it was possible to observe an increase in signal in LMB▶CD while in the LMB▶LatB condition MRTF-A was completely removed from all binding sites (Figure 3.1B). This phenomenon could be ascribed to a possible actin-dependent inhibition of MRTF-DNA binding.

To further corroborate this observation I analysed the behaviour of SRF binding at inducible sites. As published SRF binding at these sites relies on MRTF probably for nucleosome displacement and exposure of consensus sequences (Esnault et al. 2014). As shown in Figure 3.2 SRF DNA binding at the 1884 inducible sites occurred also in LMB treated cells. The SRF binding induction observed in LMB was half that observed CD (Figure 3.2B and C). Constitutive sites showed to be mostly unaffected with a mean inducibility of 0.8 and a high spearman r, confirming the linear relation between invariant binding sites as shown by the metaprofile (Figure 3.2C). I then analysed the behaviour of the SRF binding sites in LMB following the addition of either CD or LatB (Figure 3.3A). Strikingly, only the inducible sites were affected, with a 40% increase in CD and a 57% decrease in LatB (Figure 3.3A), while constitutive sites remained unaffected (Figure 3.3B). Furthermore it was possible to observe a good correlation between SRF and MRTF-A binding at inducible sites in CD, LMB and LMB▶CD conditions (Figure 3.4A). These observations were confirmed by quantitative PCR on an independent set of chromatin (Figure 3.4B).

In conclusion nuclear MRTF is sufficient for its association to genomic sites. Inactive nuclear MRTF shows reduced binding intensity and its disassociation from actin via CD enhances its association to target sites. Unexpectedly, an increase in nuclear monomeric actin using LatB impairs MRTF-DNA association although MRTF is still nuclear (see previous chapter). The association and disassociation of MRTF from target sites also affects SRF binding at inducible but not constitutive sites.

3.3 Actin-dependent DNA binding inhibition requires intact RPEL motifs

The observations described above strongly suggest that nuclear accumulation of MRTF is indeed sufficient for its association to genomic loci but nuclear actin might control it. As previously reported nuclear accumulation of MRTF via LMB preserves its interaction with nuclear actin (Vartiainen et al. 2007). Experiments collected in the lab have shown that the FRET efficiency between nuclear MRTF and actin in LMB is comparable to resting conditions. Furthermore incubation with LatB drastically enhances the MRTF-actin FRET efficiency while CD treatment completely abolishes this interaction (Vartiainen et al. 2007).

To further investigate if the effects observed at MRTF binding loci are actin mediated I compared the DNA binding activity of MRTF derivatives via ChIP. In particular I generated Doxycycline inducible cells lines expressing either MRTF fused to an NLS signal (MRTF-HA_2-NLS, FigureA) or lacking essential residues for actin binding within the RPEL motif ($MRTF^{123\text{-}1A}$-HA_2, Figure 3.5A). As shown in Figure 3.5B and C both variants are able to associate with *bona fide* SRF-MRTF targets following overnight Doxycycline induction. No detectable signal was observed in a cell line harbouring only the empty vector or expressing MRTF fused to a mutated NLS signal, confirming the specificity of the assay (Figure 3.5D and 3.6A and B). Addition of LatB completely abolished MRTF-HA_2-NLS DNA binding while $MRTF^{123\text{-}1A}$-HA_2 was only partially affected (Figure 3.5B and C and average binding in E). The reduced binding of the $MRTF^{123\text{-}1A}$-HA_2 following LatB stimulation seemed site specific as several binding sites were left completely unaffected (Figure 3.5B).

I further investigated the role of the RPEL motif in controlling MRTF-DNA association by assessing Myocardin binding behaviours. MRTF belongs to the Myocardin family of transcription co-regulators. Myocardin, like MRTF, associates to target promoters together with SRF and possesses an N-terminal RPEL domain comprising three RPEL motifs (Figure 3.7A). The actin-binding properties of Myocardin's RPEL motifs are different than MRTFs' RPELs. Specifically, MRTF's RPEL motifs show an overall greater affinity for monomeric actin than those of Myocardin (Figure 3.7 A) (Guettler et al. 2008). MRTF RPEL1 and 2 bind actin relatively strongly while MRTF RPEL3 binds actin weakly. No or weak actin-binding

is recorded for Myocardin RPEL1 and 2 while RPEL3 displays an actin-affinity comparable with that of MRTF RPEL3. RPEL3 was indeed shown to be exchangeable between the two proteins. Differences in RPEL domains define the distinct regulatory behaviours of MRTF and Myocardin. In particular while MRTFs are subject to an actin-dependent import-export regulation, Myocardin is constitutively nuclear and its activity is mostly unaffected by monomeric actin (Guettler et al. 2008). To assess if Myocardin DNA binding was affected by LatB treatments I generated Doxycycline inducible cell lines expressing either Myocardin or chimaeras of MRTF fused to either RPEL1-2 (MC-N12-MRTFA) or RPEL1-2-3 (MC-N123-MRTFA) of Myocardin (Figure 3.7B and C) (Guettler et al. 2008). Efficient DNA binding was observed for all clones at the *Cyr61* and *Acta2* promoters (Figure 3.7 D). Furthermore addition of LatB did not affect the DNA binding of Myocardin or of the MRTF variant. This observation implies that the actin-mediated MRTF DNA binding inhibition requires intact RPEL1 and 2. Chimeras of MRTF fused to the RPELs of Myocardin become resistant to the LatB effect (Figure 3.7D).

In order to elucidate if actin directly impairs MRTF-DNA association I isolated recombinant MRTF-A wild type, obtained from baculovirus-infected SF9 cells, and monomeric actin obtained in large quantities from rabbit skeletal muscle. A DNA-pull down assay was optimised in order to obtain an efficient ternary complex between SRF, MRTF and the specific DNA sequences recognised by SRF (Figure 3.8B). The addition of increasing actin quantities was shown to gradually reduce the association of MRTF with the SRF-DNA complex (Figure 3.8A). Although this preliminary data confirms that actin may directly affect MRTF-DNA association, further investigations are required including the analysis *in vitro* of MRTF mutants unable to interact with actin.

3.4 MRTF DNA binding is controlled dynamically after Rho-activation

MRTF nuclear accumulation and subsequent DNA binding can also be achieved through Rho activation (Miralles et al. 2003). This effect is caused by a drastic reduction in monomeric actin as it is assembled into filaments (Vartiainen et

al. 2007). Upon serum stimulation the kinetic of Rho activation changes over time and the formed filaments start to depolymerise.

Following these notions I wanted to see if MRTF DNA binding was controlled over time after serum-dependent Rho activation. As shown in Figure 3.9A MRTF is rapidly accumulated in the nucleus 15 minutes after serum stimulation. Afterwards its nuclear export occurs slowly still being 60-70% nuclear 1h after serum stimulation. Addition of LatB at minute 15 showed a drastic drop in overall MRTF nuclear accumulation that within 2.5 minutes becomes ~60% nuclear (Figure 3.9A). I therefore followed both MRTF and SRF binding at seven independent inducible sites where SRF binding changes with MRTF nuclear accumulation. Both MRTF and SRF bindings are strongly induced within 15 and 17.5 minutes after serum stimulation respectively (black lines in Figure 3.9B and C). MRTF and SRF binding were reduced over time showing a different kinetic than MRTF nuclear export, reaching a plateau 30 to 45 minutes after serum stimulation (Figure 3.9B and C). The LatB pulse chase experiment combined with ChIP assay showed a drastic reduction in both SRF and MRTF binding (red lines in Figure 3.9 B and C). In particular treatments with LatB for 2.5 minutes were sufficient to reduce both MRTF and SRF binding by 90 - 100% (Figure 3.9B and C) although MRTF is still 60% nuclear (Figure 3.9A). These data are consistent with observations already published where, after serum stimulation, MRTF and actin quickly disassociate 10-15 minutes after serum stimulation and re-associate showing efficient FRET at 30 minutes after serum stimulation (Vartiainen et al. 2007).

In order to investigate the relationship between MRTF DNA binding and gene activation I followed the induction of 26 target genes after serum stimulation (top 9% of the 288 genes induced specifically by CD and within 70Kb of an SRF binding site, see chapter 2). All 26 targets showed a rapid increase followed by a slow decrease in their pre-mRNA accumulation (black lines in Figure 3.9D). Addition of LatB at minute 15 was shown to drastically affect pre-mRNA accumulation that rapidly decreased within a few minutes (red lines in Figure 3.9D). LatB treatments show strong consequences in mature RNA accumulation for all 28 targets (Figure 3.9E).

I further investigated the mechanism of MRTF-dependent activation by tracking Pol II, Serine 5 and Serine 2 phosphorylation, following serum stimulation,

over the bone fide SRF-MRTF target *Acta2*. Several probes have been designed spanning the *Acta2* ORF from the upstream SRF binding site (~300 bp upstream the TSS) down to the 3' UTR region (Figure 3.10A). As shown in Figure 3.10B serum stimulation quickly induces Pol II recruitment and phosphorylation of both Ser5 and Ser2. It is intriguing to point out that at early time points after serum induction the signal of Pol II detected at the TSS is greater that the signal in the gene body (Figure 3.10B). Over the time this difference decreases and the phospho marks are sequentially reduced. I also followed the consequence of drastically removing MRTF from *Acta2* promoter via LatB treatment on Pol II recruitment and phosphorylation. Addition of LatB was shown to reduce the overall Pol II quantity at *Acta2* and immediately flatten its profile (Figure 3.10B red labelled profiles). This effect might be caused by reduced re-initiation. Furthermore LatB treatment also caused a loss of Ser2 and Ser5 apparently greater than the reduction in overall Pol II (Figure 3.10B). In conclusion the dynamics in MRTF DNA binding might have direct effects in Pol II re-initiation and phosphorylation, clearly resulting in a loss of productive transcription.

3.5 Summary

In this chapter I have presented a new role for nuclear actin in controlling MRTF DNA association. Nuclear accumulation of MRTF is sufficient for partial association of MRTF with target loci and the removal of actin using CD enhances MRTF signal. On the other hand forcing MRTF and actin interaction via LatB completely impairs MRTF-DNA association throughout the genome. Similarly SRF binding is affected at inducible sites whereas at constitutive sites SRF binding is unaffected. Furthermore inhibition of MRTF-DNA binding via LatB requires intact RPEL motifs and does not affect the related protein Myocardin. In addition MRTF DNA binding changes over time after Rho activation and addition of LatB enhances loss of MRTF binding. Changes in MRTF-DNA association correlate with the dynamic accumulation of precursor and mature RNA. Finally MRTF DNA binding seems to directly affect Pol II re-initiation and phosphorylation as shown for the *Acta2* model gene.

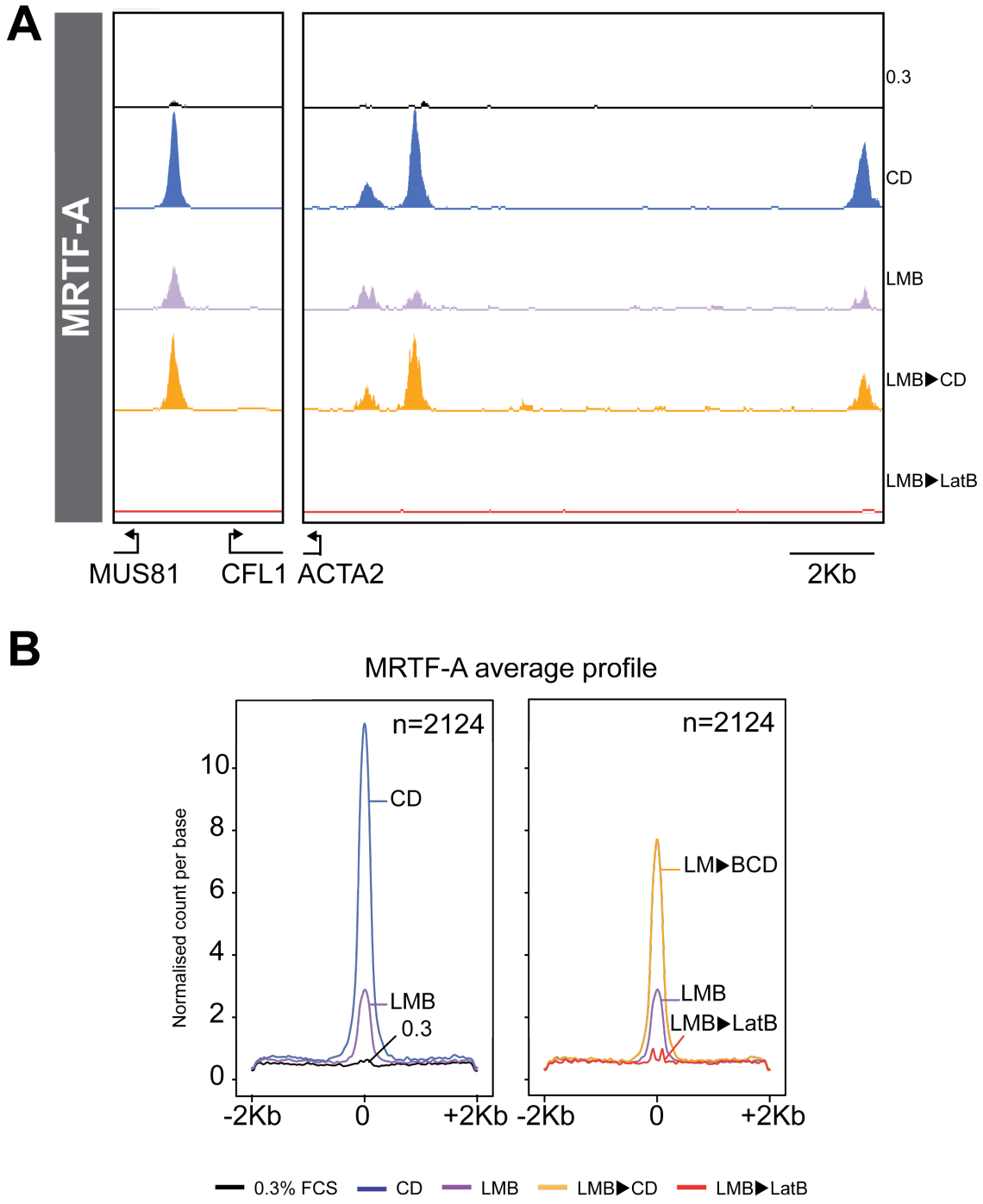


Figure 3.1 MRTF DNA binding is controlled by nuclear actin.

(A) Representative MRTF-A profiles as normalised reads per base pair at *Cofilin* (CFL1) and *Acta2* promoter. The five conditions are serum starved (0.3, black line), CD (blue line), LMB (purple line), LMB▶CD (orange line), LMB▶LatB (red line). **(B)** Metaprofile of MRTF-A binding centred on SRF peak summit.

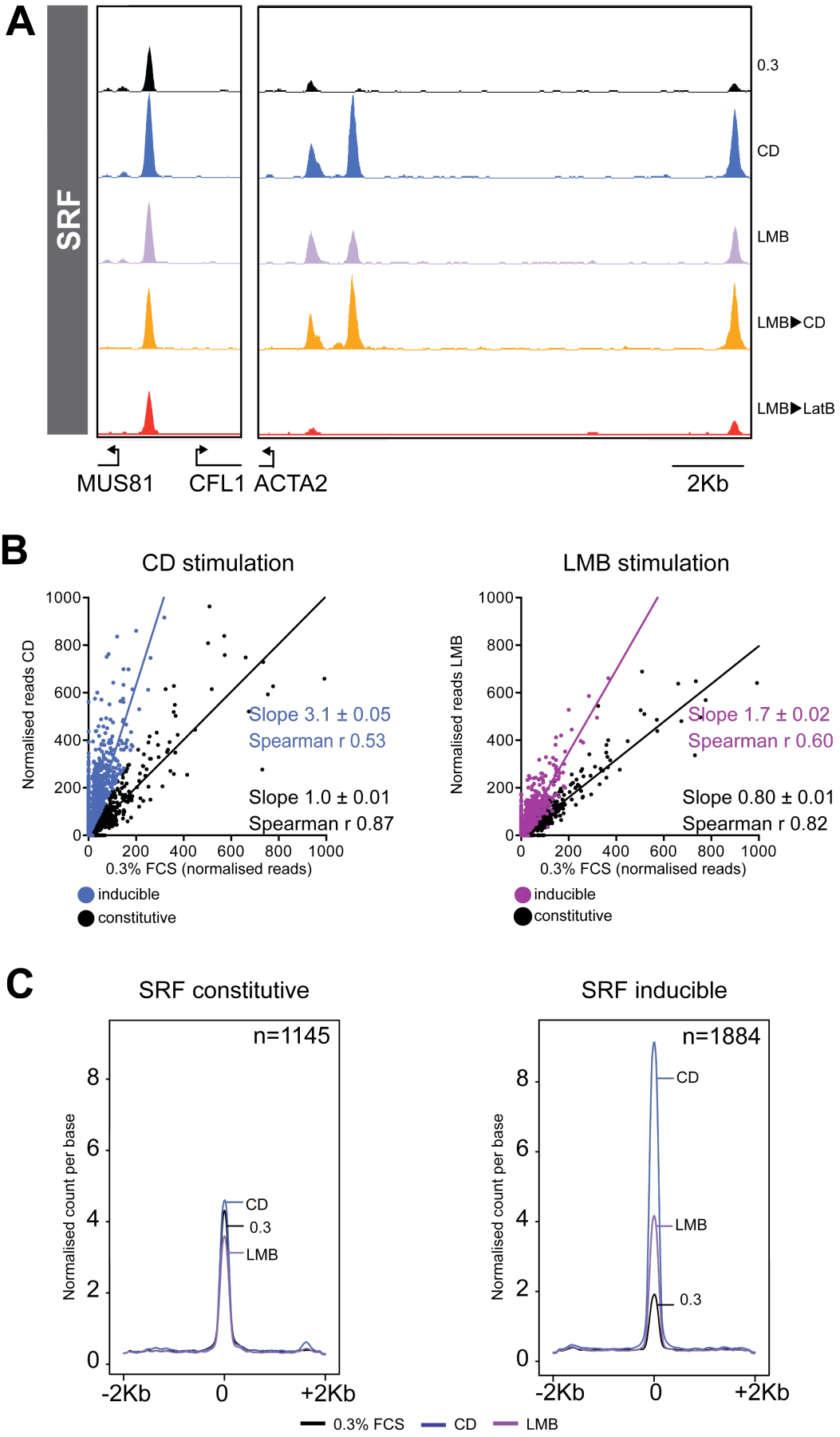

A
SRF
0.3
CD
LMB
LMB▶CD
LMB▶LatB
MUS81
CFL1
ACTA2
2Kb
B
CD stimulation
Normalised reads CD
Slope 3.1 ± 0.05
Spearman r 0.53
Slope 1.0 ± 0.01
Spearman r 0.87
0.3% FCS (normalised reads)
inducible
constitutive
LMB stimulation
Normalised reads LMB
Slope 1.7 ± 0.02
Spearman r 0.60
Slope 0.80 ± 0.01
Spearman r 0.82
0.3% FCS (normalised reads)
inducible
constitutive
C
SRF constitutive
n=1145
Normalised count per base
CD
0.3
LMB
-2Kb
0
+2Kb
SRF inducible
n=1884
CD
LMB
0.3
0.3% FCS
CD
LMB

Figure 3.2 MRTF nuclear accumulation is sufficient to induce SRF binding.

(A) Representative SRF profiles as normalised reads per base pair at *Cofilin* (CFL1) and *Acta2* promoter. **(B)** MRTF nuclear accumulation via CD or LMB enhances SRF binding at inducible sites. Scatter plot comparing ChIP-seq read counts in CD (right) or LMB (left) stimulation and resting cells. Each graph is divided in constitutive sites (<1.9 fold increase black) and inducible (>1.9 fold increase, blue for CD stimulation left panel or purple for LMB stimulation right panel). Solid lines show linear regression plots for the two populations. **(C)** Metaprofiles of SRF binding at constitutive (*left*) and inducible (*right*) sites.

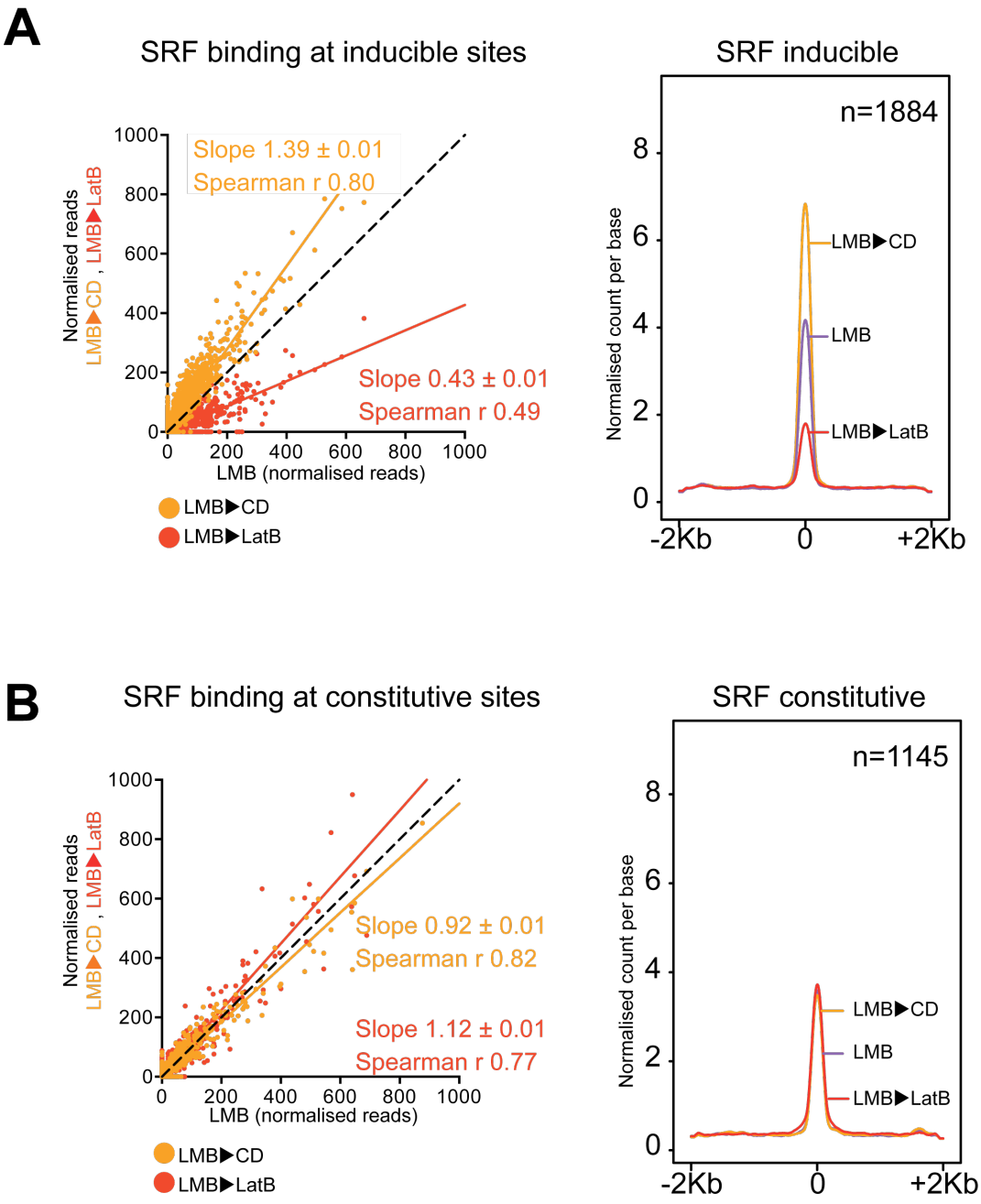


Figure 3.3 Nuclear actin controls SRF inducible binding via MRTF.

(A) (*Left*) Scatter plot of SRF binding at inducible sites comparing ChIP-seq read counts in LMB▶CD (orange) or LMB▶LatB (red) and LMB stimulated cells. (Dotted line) bisector of the scatter plot distribution representing invariant binding sites compared to LMB stimulation. Solid lines shows linear regression plot for each condition. (*Right*) Metaprofile of SRF binding in LMB, LMB▶CD and LMB▶LatB at inducible sites (n=1884). **(B)** (*Left*) Scatter plot of SRF binding at constitutive sites comparing ChIP-seq read counts in LMB▶CD (orange) or LMB▶LatB (red) and LMB stimulated cells. (Dotted line) bisector of the scatter plot distribution representing invariant binding sites compared to LMB stimulation. Solid lines shows linear regression plot for each condition. (*Right*) Metaprofile of SRF binding in LMB, LMB▶CD and LMB▶LatB at constitutive sites (n=1145).

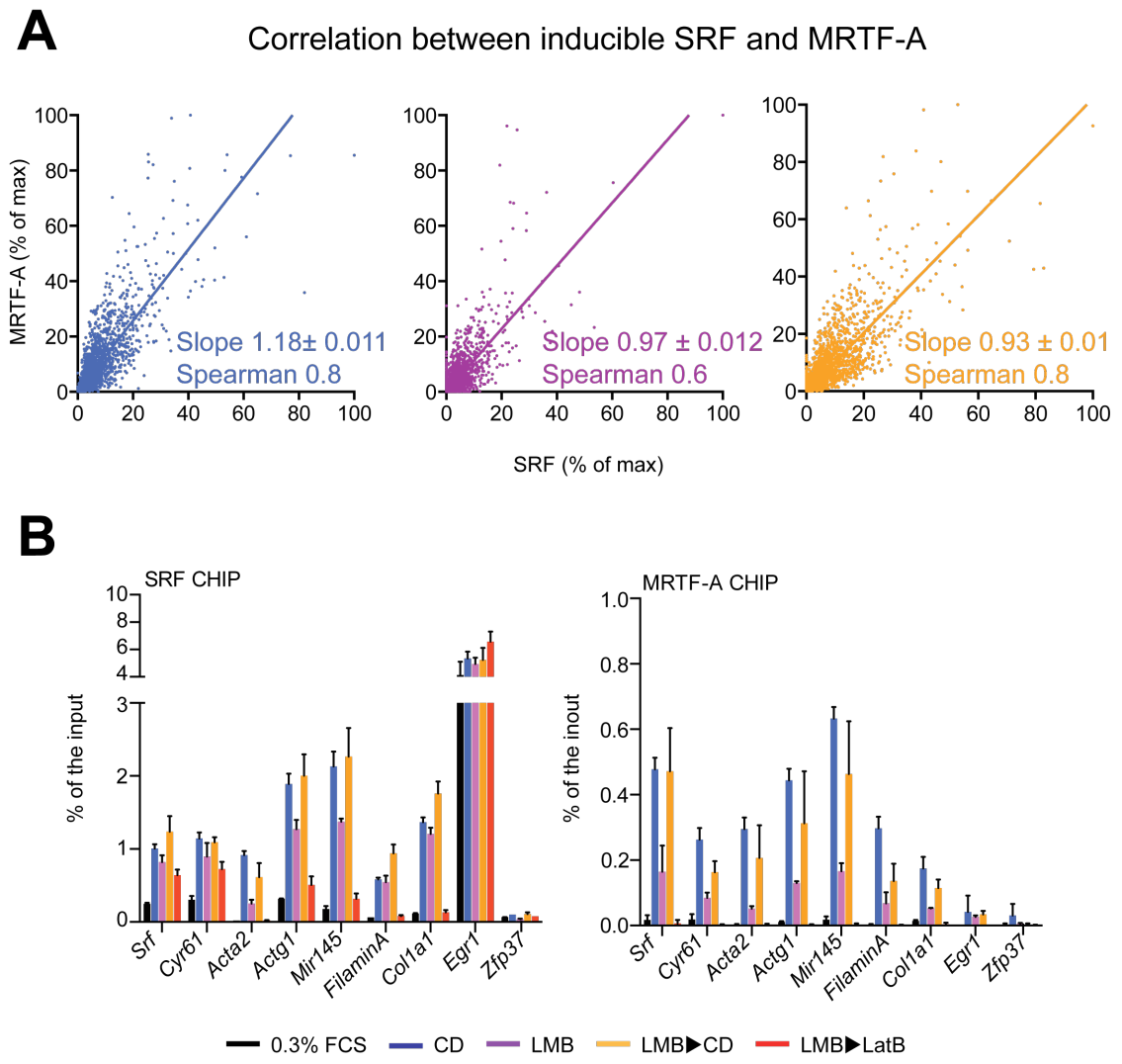


Figure 3.4 SRF and MRTF binding correlates at inducible sites.

(A) Correlation between SRF and MRTF ChIP-seq reads at inducible sites. Both SRF and MRTF normalised ChIP-seq read counts were scaled to the highest signal across condition (with maximum at 100%). The scatter plot represents correlation in CD (blue), LMB (purple) and LMB▶CD (orange) conditions. Solid line shows linear regression plot for each condition. **(B)** ChIP validation of SRF binding (*left*) and MRTF binding (*right*) at nine sites including: two SRF constitutive MRTF bound sites (*Srf* and *Cyr61*), five SRF inducible MRTF associated sites (*Acta2, Actg1, Mir145, Filamin A* and *Col1a1*), one SRF constitutive TCF specific site (*Egr1*) and one negative control site where both SRF and MRTF are not associated (*Zfp37*). The IP was quantified using quantitative RT-PCR and plotted as percentage of the input material.

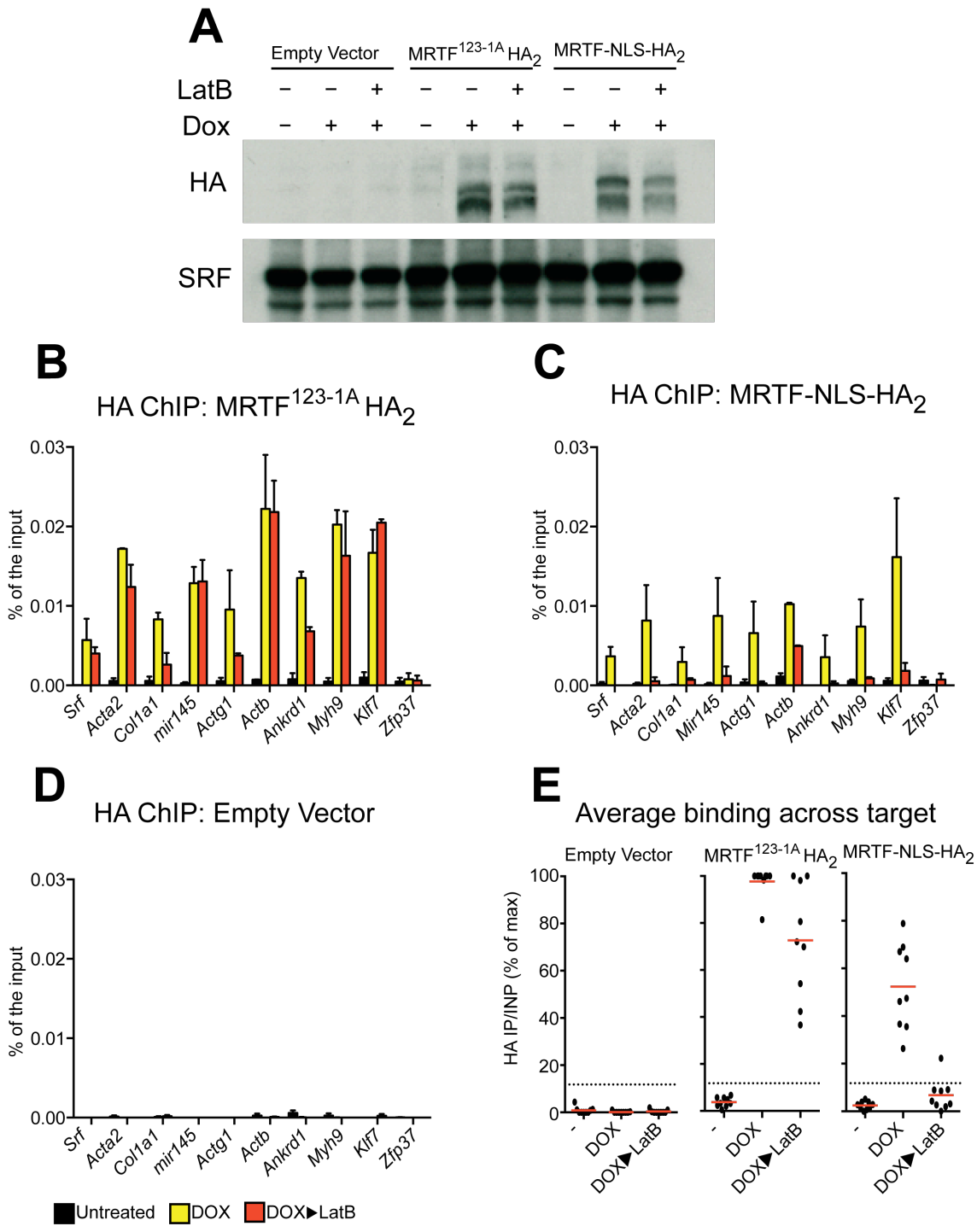

A
Empty Vector
MRTF123-1A HA2
MRTF-NLS-HA2
LatB
Dox
HA
SRF
B
HA ChIP: MRTF123-1A HA2
% of the input
Srf
Acta2
Col1a1
mir145
Actg1
Actb
Ankrd1
Myh9
Klf7
Zfp37
C
HA ChIP: MRTF-NLS-HA2
Mir145
D
HA ChIP: Empty Vector
Untreated
DOX
DOX▶LatB
E
Average binding across target
HA IP/INP (% of max)
DOX
DOX▶LatB

Figure 3.5 Actin mediated MRTF DNA binding inhibition requires intact RPEL motifs.

(A) Western blot analysis of cell lines carrying Empty vector, $MRTF^{123\text{-}1A}HA_2$, MRTF-NLS-HA_2. Cells were stimulated with 2µg/ml doxycycline (Dox) and treated with LatB for 5 minutes to minimise unspecific LatB effects. **(B to D)** ChIP using an anti-HA antibody of MRTF-NLS-HA_2, $MRTF^{123\text{-}1A}HA_2$ and cells harbouring the empty vector. The immune-precipitated material was quantified using quantitative RT-PCR and plotted as percentage of the input material. The binding sites assessed include nine *bona fide* SRF-MRTF targets and a negative control (zfp37) **(E)** Average binding across targets and cell lines. The signal across cell lines per binding site was normalised to the highest signal measured.

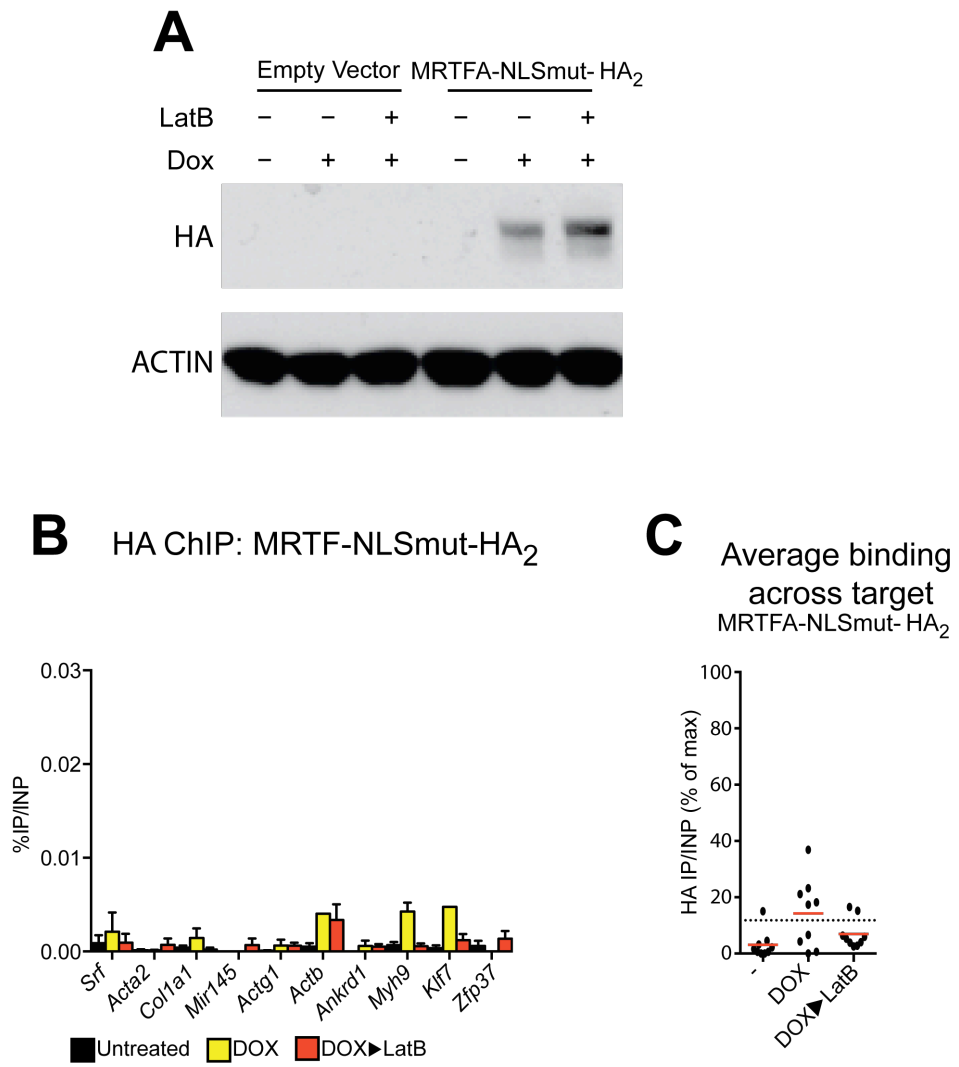


Figure 3.6 Expression of MRTF linked to a mutated NLS does not induce DNA binding.

Control experiment related to figure 3.5. **(A)** Western blot analysis of cell lines carrying Empty vector or MRTF-NLSmut-HA_2. Cells were stimulated with 2µg/ml doxycycline (Dox) and treated with LatB for 5 minutes to minimise unspecific LatB effects. **(B)** ChIP using an anti-HA antibody of MRTF-NLSmut-HA_2 **(C)** Average binding across targets and cell lines. The signal per binding site was normalised to the maximum signal measured including the experiment seen in Figure 3.5.

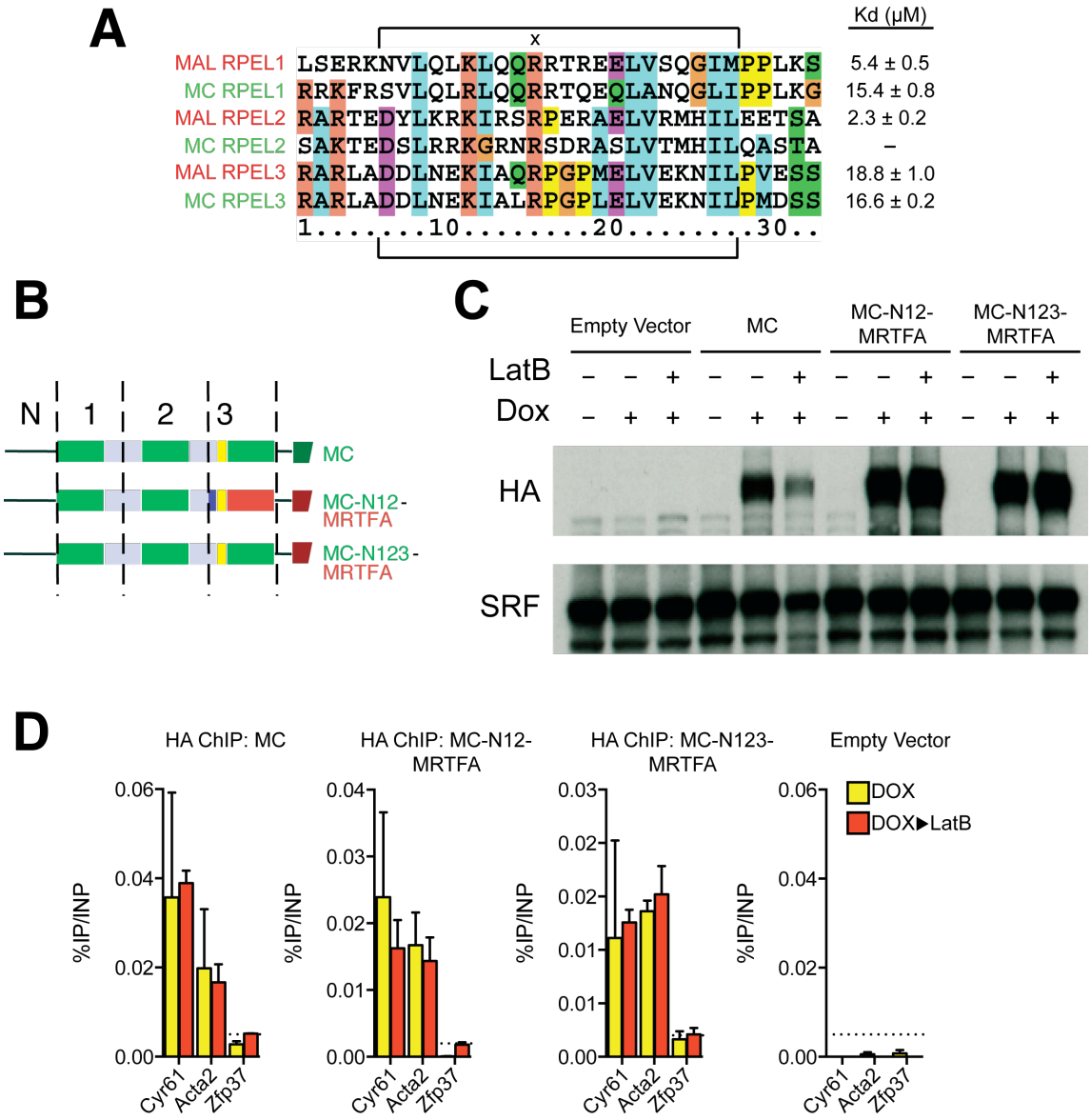


Figure 3.7 Actin-mediated MRTF DNA binding inhibition requires RPEL1-2.

(A) (*Left*) Multiple sequence alignment of RPEL motifs of mouse MRTF and MC. This figure was modified from (Guettler et al. 2008). The brackets indicate the RPEL motif as defined by PFAM. "x" denotes the first most conserved R residue of the RPEL motif targeted by the R to A mutation in $MRTF^{123\text{-}1A}HA_2$. (*Right*) Kd of each RPEL motif for monomeric actin is listed (Guettler et al. 2008). **(B)** Schematic representation of the MC and MC-MRTF chimeras with MC sequences in green and grey and MRTF sequences in red and blue (panel adapted from Sebastian Guettler's thesis). **(C)** Western blot analysis using an anti-HA antibody of cell lines carrying Empty vector, MC, MC-N12-MRTFA and MC-N123-MRTFA. Cells were stimulated 2μg/ml doxycycline (Dox) and treated with LatB. **(D)** Anti-HA ChIP of the MC, MC-N12-MRTFA, MC-N123-MRTFA and control cell line. The immune-precipitated material was quantified using quantitative RT-PCR and plotted as percentage of the input material. The binding sites assessed include two *bona fide* SRF-MRTF targets and a negative control (*Zfp37*).

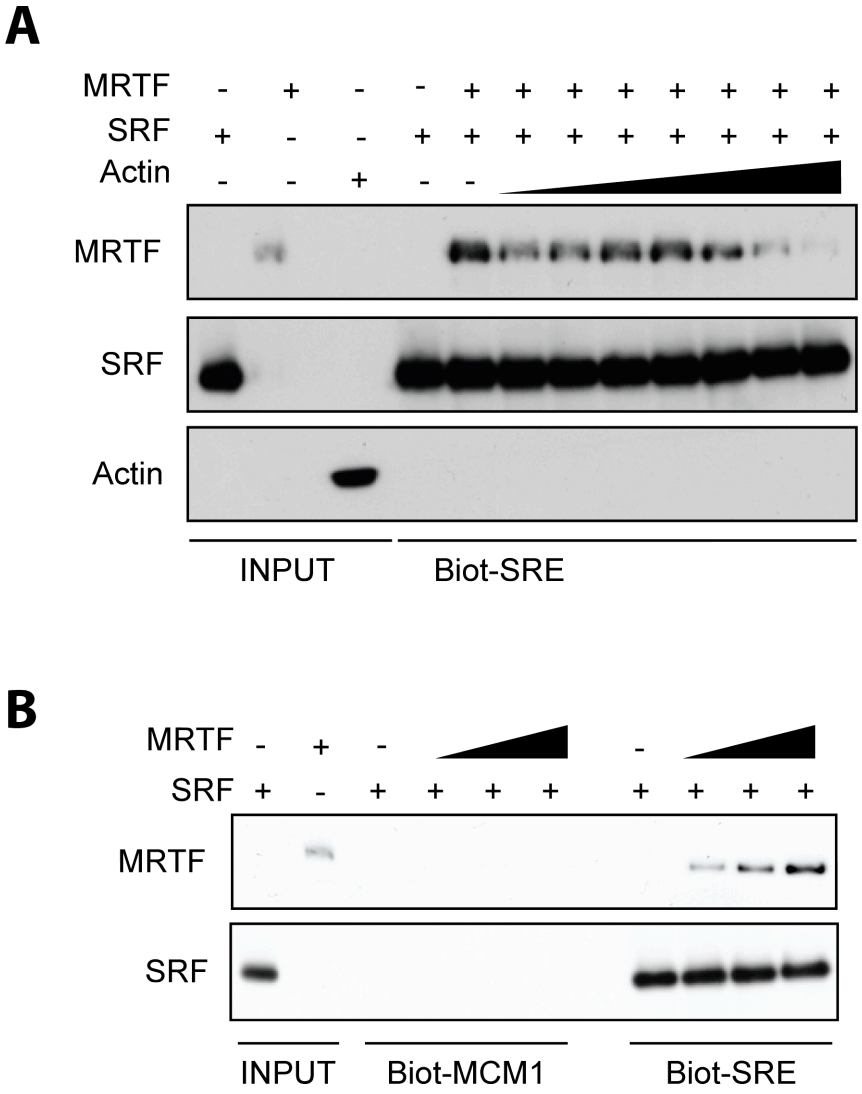


Figure 3.8 Monomeric actin directly affects MRTF DNA binding *in vitro*.

(A) DNA-protein binding assay performed using an SRF-MRTF specific binding site. The DNA-protein complex was pulled down using Dynabeads M-280 Streptavidin. SRF and MRTF pure components were coupled to the DNA. After extensive washes increasing amounts of pure monomeric LatB:actin were added to the DNA-SRF-MRTF complex and incubated for 15 minutes at 30 ºC. **(B)** Optimisation of the DNA-pull down conditions using a biotinylated DNA probe with (Biot-SRE) or without (Biot-MCM1) the specific SRF binding site.

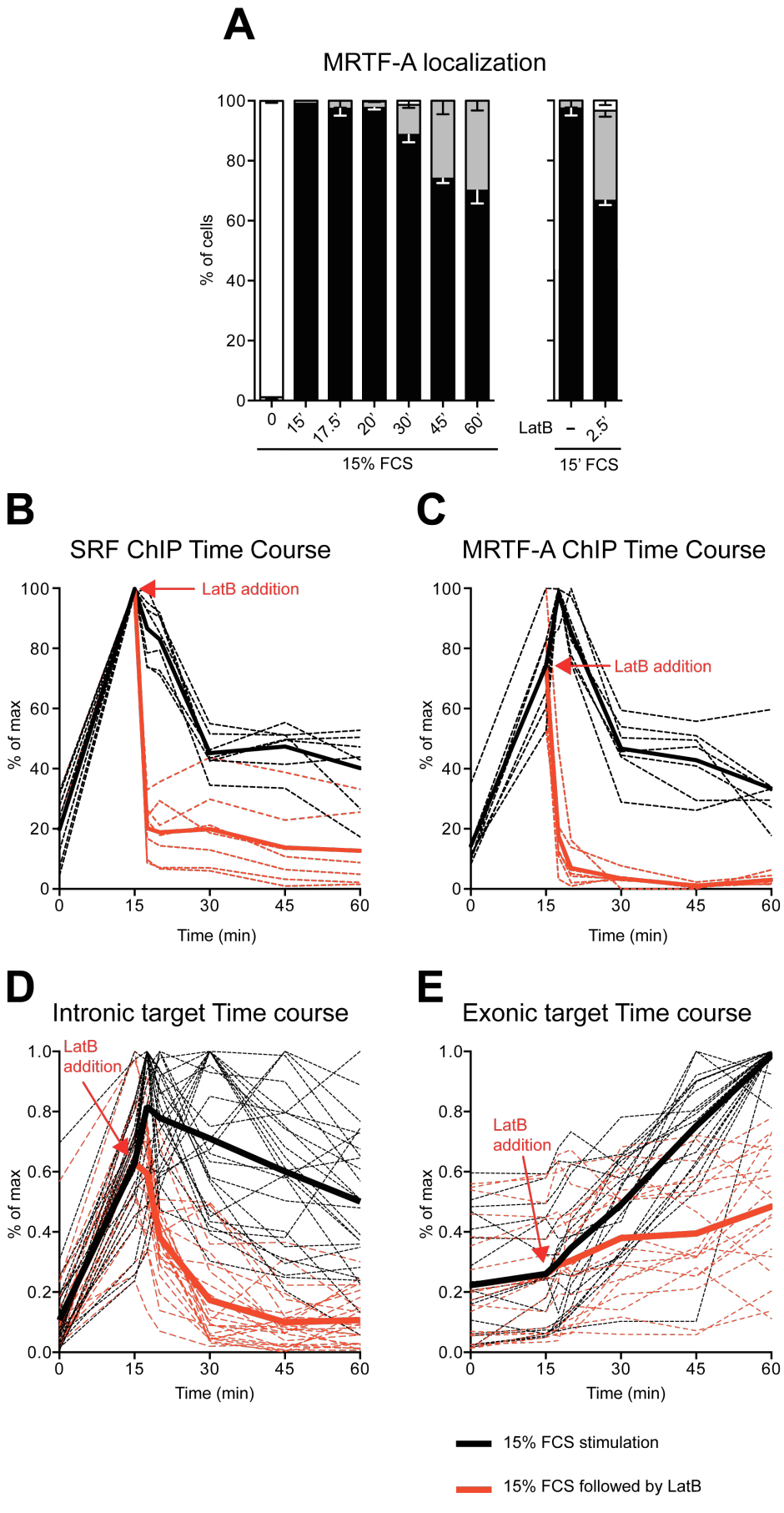

A
MRTF-A localization
% of cells
0
20
40
60
80
100
0
15'
17.5'
20'
30'
45'
60'
15% FCS
LatB
–
2.5'
15' FCS
B
SRF ChIP Time Course
LatB addition
% of max
Time (min)
C
MRTF-A ChIP Time Course
LatB addition
% of max
Time (min)
D
Intronic target Time course
LatB addition
% of max
Time (min)
E
Exonic target Time course
LatB addition
% of max
Time (min)
15% FCS stimulation
15% FCS followed by LatB

Figure 3.9 MRTF DNA binding changes after Rho activation.

(A) (*Left*) MRTF-A nuclear localisation time course after serum stimulation of serum starved cell. 150-300 cell were scored according to predominantly nuclear (N), pancellular (N/C) and predominantly cytoplasmic (C) MRTF-A localisation. (*Right*) MRTF-A nuclear localisation following a quick pulse of LatB (2.5 minutes). **(B and C)** SRF (*left*) and MRTF (*right*) ChIP time course at seven inducible binding sites. The signal per binding site was normalised to the highest point across the time course. The Black dotted lines represent each independent binding site while the solid black line represent the average binding intensity per time point after serum stimulation. The red dotted lines represent each independent binding site while the solid red line represent the average binding intensity per time point after LatB pulse chase following 15 minutes of serum stimulation. **(D)** Time course expression of 26 independent SRF-MRTF target genes following accumulation of their first intron after serum stimulation (Black line) or LatB pulse chase after 15' serum shock (red line). Dotted lines represent individual genes while the solid lines represent an average per time point across the 26 genes. Quantitative PCR were performed for each gene and the values were normalised to Gapdh. Each kinetic was normalised to the highest point across the time course. **(E)** As in D but following accumulation of mature RNA of the 26 targets analysed in panel D. (The 26 targets are: *Ctgf, Vcl, Srf, Acta2, Dusp5, Pdlim5, Cyr61, Serpine1, Tpm1, Thbs1, Palld, Bok, Ankrd1, Wdr1, D4Bwg0951e, Pdlim7, Dstn, Vgll3, Slc2a1, Klf6, Rheb, Tpm4, Klf7, Zyx, Pdcl3 and Actb*).

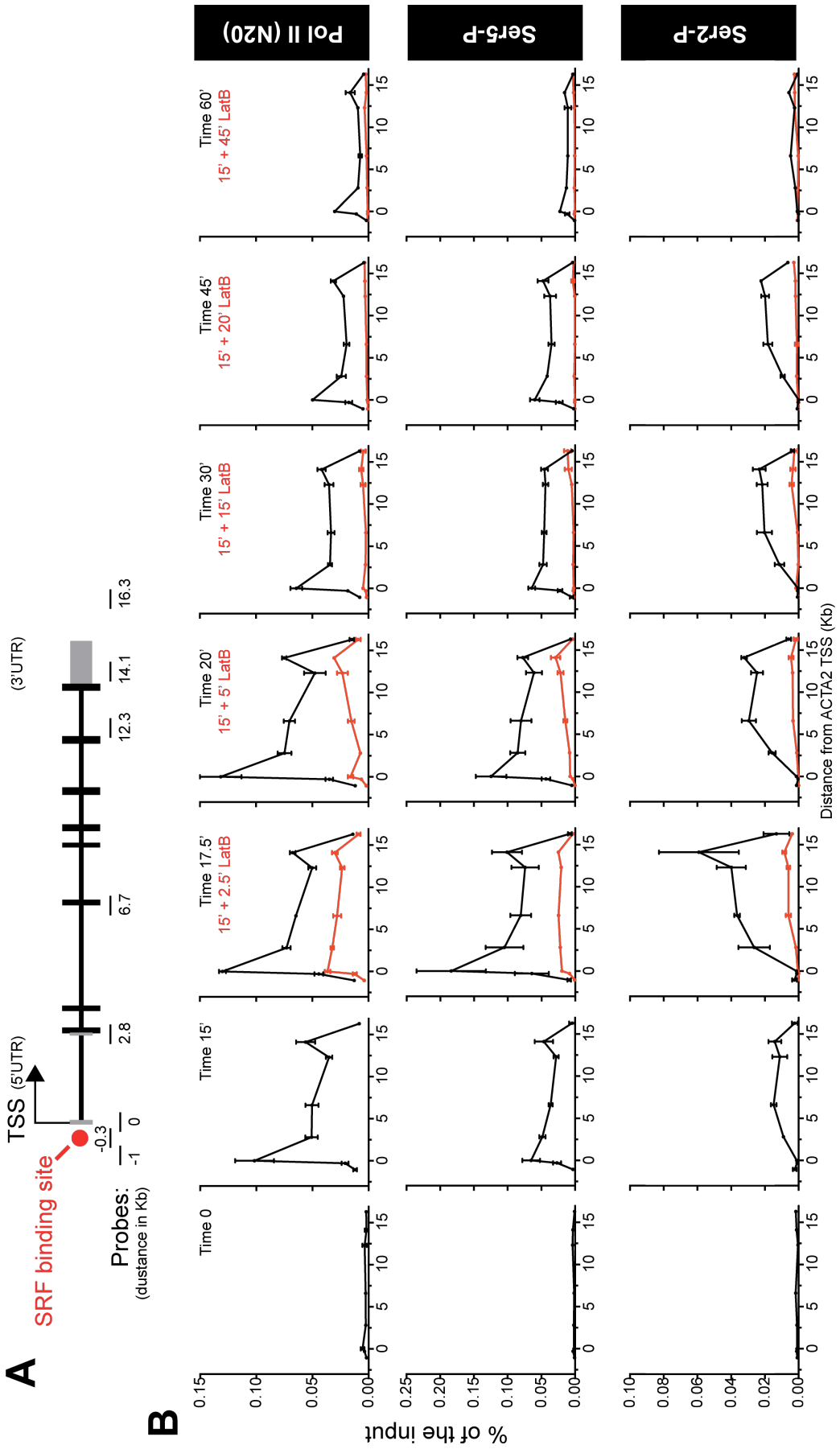

A
SRF binding site
TSS (5'UTR)
(3'UTR)
Probes:
(dustance in Kb)
-0.3
-1
0
2.8
6.7
12.3
14.1
16.3
B
Pol II (N20)
Ser5-P
Ser2-P
Time 0
Time 15'
Time 17.5'
15' + 2.5' LatB
Time 20'
15' + 5' LatB
Time 30'
15' + 15' LatB
Time 45'
15' + 20' LatB
Time 60'
15' + 45' LatB
% of the input
Distance from ACTA2 TSS (Kb)

Figure 3.10 MRTF DNA binding promotes Pol II recruitment and phosphorylation after serum stimulation.

(A) Representation of the *Acta2* gene illustrating the positions of different probes along the gene. Exons are indicated with 5'UTR and 3'UTR in grey and coding sequences in black; SRF binding site is indicated with a red circle. **(B)** ChIP of Pol II (N20 antibody), Serine 5-P (H14 antibody) and Serine 2-P (H5 antibody) over the time course illustrated in Figure 3.9. Black lines represent the profile of Pol II and the phosphorylated CTD forms after serum stimulation. In red is the profile following LatB pulse chase added 15 minutes after the serum shock.

Chapter 4. Nuclear actin controls Pol II productive transcription

4.1 Aim

Activation of transcription follows defined sequential events. Pol II recruitment together with the GTFs is the first mandatory bottleneck to go through in order to induce transcription (see Chapter 1 for a full description). Following Pol II recruitment other barriers need to be overcome including recruitment of CTD kinases, release from pausing and elongation. These processes are directly influenced by a plethora of effectors either promoting or inhibiting the transcriptional flow. The definition of each step is often misleading, as the exact contribution of each transcriptional factor is not yet understood. Furthermore several PTMs decorate Pol II CTD as it crosses each step. However, their relation with escape, elongation and the synthesis of mRNA is not clear.

Accumulation of MRTF into the nucleus after Rho activation influences both Pol II recruitment and activation (Esnault et al. 2014). The association of MRTF to defined genomic loci could potentially affect the transcriptional outcome directly recruiting transcriptional kinases, promoting chromatin modification and favouring nucleosome clearance at selected TSS. To further investigate how MRTF influences transcription I analysed the steps in transcriptional activation under the productive and unproductive conditions described in chapter 2 and 3 (CD and LMB stimulation). In this chapter I am going to describe how nuclear actin impacts on the transcriptional outcome affecting steps beyond Pol II recruitment and escape.

4.2 MRTF nuclear accumulation is sufficient for Pol II recruitment and escape

As described in Chapter 2 and 3 MRTF nuclear accumulation via LMB is sufficient for partial DNA binding but productive transcription requires disassociation from nuclear actin. Furthermore, as mentioned at the end of the previous Chapter and as recently published, MRTF-DNA binding induces Pol II recruitment and activation (Esnault et al. 2014). To further understand how nuclear actin negatively regulates productive transcription of MRTF-target genes I

assessed the recruitment of Pol II at the *Acta2* gene in the conditions previously described (Figure 4.1A and B). Surprisingly Pol II showed to be efficiently recruited and released from *Acta2* promoter after LMB treatment (Figure 4.1B). Pol II signal shows to be only partially reduced if compared to the productive conditions CD and LMB▶CD (Figure 4.1B). Furthermore both NELF-A and SPT5 were shown to be recruited together with Pol II and, consistent with efficient Pol II release, only SPT5 showed a further distribution within *Acta2* ORF (Figure 4.1 B). This observation contradicts to the complete loss of efficient precursor and messenger RNA accumulation after LMB stimulation as described with the RNA-seq analysis (Figure 4.1C).

In order to validate this observation I performed a genome-wide analysis of the Pol II ChIP (Figure 4.2A). I considered the specific set of genes induced by CD and unaffected by LMB within 70kb of an SRF binding site as described in Chapter 2. It was possible to observe an increase in Pol II signal across all 288 MRTF-specific targets in both CD and LMB (Figure 4.2B and D). Furthermore the LMB-dependent induction of Pol II recruitment showed a marked sensitivity to LatB treatment (Figure 4.2B). This increase in Pol II occurred identically at MRTF-specific genes where LMB does not induce any transcription and at genes where LMB just mildly enhances the intronic-reads base line (see Chapter 2 for a definition of the gene group, Figure 4.2C). No change was observed at constitutively active genes while strong recruitment of Pol II, travelling within the gene body, was specifically recorded for the 288 MRTF-specific target genes (Figure 4.2D). Overall the Pol II signal in LMB was half that of observed in CD but no difference was observed for the LMB▶CD condition (Figure 4.2D).

In conclusion, in the absence of productive transcription nuclear MRTF is sufficient to recruit and release Pol II from MRTF-specific target genes. The detected Pol II specifically requires MRTF DNA binding as treatment with LatB induces Pol II loss. This change in Pol II distribution is specific for MRTF-targets, as it is not observed at any constitutively active genes.

4.3 Pol II recruitment correlates with MRTF binding but actin disassociation is required for RNA-synthesis

In order to asses whether MRTF binding correlates with Pol II recruitment I compared the distribution of the Pol II signal at the TSS of each MRTF-specific target and the highest, closest peak of MRTF (Figure 4.3). Although it was challenging to establish a unique relation between MRTF-binding sites and target genes, it was still possible to record a correlation in CD, LMB and LMB▶CD treatments (Spearmen r=0.3 P-value <0.0001). LatB treatments showed no correlation between MRTF and Pol II signal (Figure 4.3).

4.4 Pol II escape is accompanied by RNA synthesis and R-loop accumulation

The relationship between travelling Pol II and the production of RNA has been shown to be linear (Tippmann et al. 2012). Also in my assays, the comparison of Pol II signal and RNA synthesis after CD stimulation showed a good linear correlation with a slope of 0.85 and Spearman r of 0.4 (P-value<0.0001) (Figure 4.4A). The comparison of Pol II signal and RNA synthesis in LMB treated samples showed a distribution with slope of 0.18 and Spearman r 0.2 (p-value<0.0001) implying that most increases in Pol II signal are not followed by a proportional increase in RNA synthesis (Figure 4.4B). A proportional stimulation of Pol II signal and RNA synthesis is re-established by addition of CD (Figure 4.4C left panel), while LatB treatment reduces the Pol II signal observed in LMB (Figure 4.4C right panel).

In order to asses if the engaged unproductive Pol II is transcriptionally competent I performed Run-on assays after short CD or LMB stimulations. Accumulation of labelled RNA was detected in both CD and LMB condition across the *Acta2* gene (Figure 4.5 A left panel). Furthermore the amount of RNA detected in LMB could easily be related to the amount of Pol II seen by Pol II ChIP over *Acta2* ORF (Figure 4.1B). The use of conventional RNA extraction methods, following a short time course after either CD or LMB stimulation, did not show any accumulation of *Acta2* precursor in LMB treatment compared to CD (Figure 4.5A).

I further studied the ability of the engaged Pol II to synthesise RNA, assessing R-loop accumulation. R-loops are three stranded nucleic acid structures formed by RNA:DNA hybrids plus a displaced DNA strand (ssDNA) (Aguilera & García-Muse 2012). This structure is a transcriptional by-product. The RNA associated with the DNA strand is a fairly stable hybrid that could be detected using specific antibodies (Boguslawski et al. 1986). Similarly to the observed Pol II distribution, R-loops are accumulated following Pol II elongation in CD, LMB and LMB▶CD conditions over *Actb* and *Acta2* (Figure 4.6 B top two panels) and the detected signal showed sensitivity to RNAseH treatments (Figure 4.6 B bottom two panels).

These observations suggest that MRTF-DNA binding is sufficient for Pol II elongation but actin disassociation is required so that the synthesised RNA can be efficiently accumulated. In addition these results confirm that the phenomenon observed following LMB stimulation is not caused by an impairment of Pol II enzymatic activity.

4.5 Nuclear actin affects Pol II phosphorylation

To further understand which step in transcription is affected by nuclear actin I assessed the phosphorylation of Pol II CTD in the panel of conditions analysed so far. As extensively described in the introduction (see Chapter 1) Pol II CTD is phosphorylated throughout the transcriptional cycle. Five key residues have shown to be phosphorylated by diverse kinases and their function in yeast, drosophila and mammals have been partially described (Figure 4.6A). The pattern of phosphorylation has been correlated with some transitions that Pol II has to cross in order to efficiently transcribe. Furthermore several complexes have been shown to specifically interact with phosphorylated Pol II CTD and directly affect steps in RNA synthesis, maturation, release and export.

Using the *Acta2* model gene I compared the signal distribution of five phospho-specific antibodies. Induction of productive transcription via CD induces phosphorylation at all residues (Figure 4.6B to F). As described by others, both Serine 5 and 7 are enhanced at the *Acta2* TSS (Figure 4.6B and C) and spread down towards the end of the gene following the Pol II distribution (Figure 4.6 G). Serine 2, on the other hand, showed a bell shaped distribution starting to

accumulate immediately after the TSS and peaking towards the 3' UTR (Figure 4.6D). Tyr1, one of the least described markers together with Thr4, shows high signal at the *Acta2* promoter and a small spike after the 3' UTR (Figure 4.6 E). On the contrary Thr4, as recently described (Hintermair et al. 2012), is accumulated immediately after the termination site at *Acta2* 3' UTR (Figure 4.6 F). LMB treatment showed a different scenario than CD stimulation. Only Serine 5 phosphorylation accompanied the engaged Pol II from the promoter down to the 3' UTR (Figure 4.6 B and G). Serine 5 signal in LMB showed LatB sensitivity again suggesting the direct requirement of MRTF bound at the Acta2 promoter. On the other hand Serine 7, Serine 2 Tyrosine 1 and Threonine 4 were all defective (Figure 4.6 C to F). Stimulation with CD following LMB was shown to enhance only a few phospho-marks. In particular Thr4 was still defective showing only a barely detectable signal down the 3' UTR.

In order to generalise these observations I performed a genome-wide analysis of both Serine 5 and Serine 2 ChIP (Figure 4.7). As shown in Figure 4.8A and C LMB stimulation induces Serine 5 accumulation at all 288 MRTF-target genes. This stimulation similarly occurred at MRTF-specific genes where LMB has no effect and at genes where LMB mildly enhances the base line of intronic-reads (Figure 4.8 A). Addition of CD partially enhances the Ser5-P signal following LMB stimulation while LatB treatment specifically impairs the LMB-induced Ser5-P signal (Figure 4.8 A and C). No changes in Serine 5 phosphorylation were observed at active uninduced genes (Figure 4.8 C). Analysis of Serine 2 phosphorylation over the MRTF-specific genes showed a different scenario. Complete loss of phospho-Serine2 in LMB treatment occurred at all MRTF-specific genes (Figure 4.8 B and D). No difference was recorded for genes where LMB partially enhances the RNA base line (Figure 4.8B). Addition of CD following LMB partially enhances Serine 2 phosphorylation (Figure 4.8B and D).

In conclusion the synthesis of stable RNA reflects a correct phosphorylation of Pol II CTD. During unproductive stimulations, even though Pol II is engaged and released, the phospho-signature at Pol II CTD is affected. Further studies are going to be essential to fully characterise this unproductive condition including genome-wide analysis of the remaining three CTD modifications. However it is clear that nuclear actin via MRTF affects steps beyond Pol II recruitment and escape. In particular, given the effect of actin on MRTF-DNA binding, it would be important to

study how a TF or co-regulator such as MRTF can affect steps in transcription by changing its persistence at target promoters.

4.6 MRTF nuclear accumulation does not influence Pol II escape

To further characterise the process of MRTF-dependent transcription activation, given the data collected, I analysed the distribution of Pol II signal in the conditions used. As reported by the Adelman and Lis groups Pol II promoter recruitment is followed by a step of pausing within a few hundred base pairs from the TSS (Adelman & Lis 2012). This mechanism relies on the so-called pausing factors NELF and DSIF that play a dual role in PIC formation and Pol II release repression. Several studies described how different TFs are capable of recruiting p-TEFb, a complex required for pause release induction (Rahl et al. 2010). As previously described within this chapter MRTF nuclear accumulation allows Pol II recruitment and release with a correct redistribution of SPT5 and NELF eviction suggesting that, even though no RNA is being detected, p-TEFb might be correctly recruited.

During the process of gene activation two main steps seem to be the bottleneck in productive elongation: Pol II recruitment and escape. Several works have tried to characterise which of these two is the rate-limiting step able to predict the degree of activity of a given gene. The ratio between the signal at the 5' end of a gene and the one observed in the gene body (GB) provides a way to measure the rate of entry and escape for a given gene. I defined the elongation index as the ratio between the GB density and the 5' region density (Figure 4.9 A). This value is the reverse of the so-called Travelling ratio or Pausing index described by others (Reppas et al. 2006; Rahl et al. 2010; Zeitlinger et al. 2007).

The relative rates of Pol II recruitment and escape combine to determine the level of transcription of a given gene (Core & Lis 2008). I therefore analysed the relation between 5' and GB signals for constitutively active genes and for the selected set of MRTF-specific genes in resting, CD and LMB conditions (Figure 4.10 B). In particular I compared the top 20% of constitutively expressed genes with the 288 MRTF-dependent targets. On average, across conditions, the 5' region shows a signal that was ~60% higher than the GB region in constitutive

active genes (Figure 4.9B red populations). Within this set of genes this difference did not change in any condition. I then focused on the 288 targets. As shown in Figure 4.9B the 5' to GB ratio did not change when the 288 genes were stimulated with either CD or LMB. This observation would imply that the elongation index per gene does not change when these genes are specifically and rapidly turned on. Indeed it was possible to observe a good relation between elongation indexes in resting and stimulated conditions for both constitutively active and MRTF-specific gene sets (Figure 4.9C). Furthermore LMB stimulation was not shown to differ from the CD stimulation.

Analysis of the changes observed at 5' and GB regions (as fold over 0.3% FCS) showed a proportional increase upon stimulation of both values (Figure 4.10). No increase was detected at highly expressed-constitutive genes while the MRTF-dependent targets showed, in both CD and LMB, an overall equal induction if compared to 0.3% FCS of the 5' Pol II peak and the GB signal.

Taking together these observations it is possible to conclude that MRTF-DNA binding primarily controls Pol II recruitment. The observation that the elongation index is maintained unchanged across conditions, suggests that the extent of Pol II escape may be modulated by mechanisms intrinsic to the transcription machinery. Corroborating this hypothesis is the fact that the rate of Pol II escape does not change even in LMB, where no accumulation of stable messenger RNA is detected.

4.7 Pause Pol II inversely correlates with SRF binding and gene induction

As no differences in Pol II escape are measured following MRTF nuclear accumulation, I assessed whether the rate of escape could be defined on a gene-by-gene bases as a function of the transcription induction. I therefore analysed both elongation and pausing indexes in 0.3% FCS. The 288-gene set of targets was ranked on the basis of the change in gene expression observed following CD stimulation (Figure 4.11A). Considering the top 80% induced genes (top four quintiles) it was possible to see that highly induced genes had a higher elongation rate (and a lower pausing index) than less induced genes in resting conditions (Figure 4.11B). As described in the previous section the rates did not change upon

stimulation with either CD or LMB. This implies that a gene, in order to be highly induced, should have lower paused Pol II to start. In addition it was possible to observe that lowly induced genes showed a higher pausing-index (Figure 4.11B) and also a higher base-line expression (Figure 4.11C). Activation of transcription via CD induces accumulation of RNA-seq intronic reads in all quintiles, but the top 20% induced genes showed the greatest change (Figure 4.11C). These changes were not observed in LMB as already discussed in Chapter 2.

Taking together these observations it is possible to conclude that the rate of Pol II escape is context specific and does not change when MRTF is retained in the nucleus. It is important to bear in mind that for the MRTF-specific set of targets there is a direct relation between inducibility and SRF proximity (see Chapter 2 and Figure 4.11D). It is therefore conceivable to suggest that for SRF-distal genes other TFs favour Pol II initiation and MRTF contributes to enhance its rate. On the other hand, for SRF-proximal genes MRTF is the primary source of initiation, therefore no Pol II and base-line expression is detected.

4.8 Summary

In this Chapter I characterised how MRTF-dependent genes respond following MRTF nuclear accumulation. I analysed the distribution of components of the PIC, including pausing factors, and modifications of the Pol II CTD in productive and un-productive conditions. MRTF DNA binding directly affects Pol II initiation. MRTF-actin disassociation is essential for the correct phosphorylation of Pol II CTD. Indeed nuclear accumulation of MRTF via LMB is sufficient for Pol II entry and escape in absence of transcript accumulation. These observations allowed me to verify that Pol II escape can be in part disentangled from its phosphorylation. A further characterisation of the modifications at Pol II CTD in these conditions will provide further insight into this mechanism.

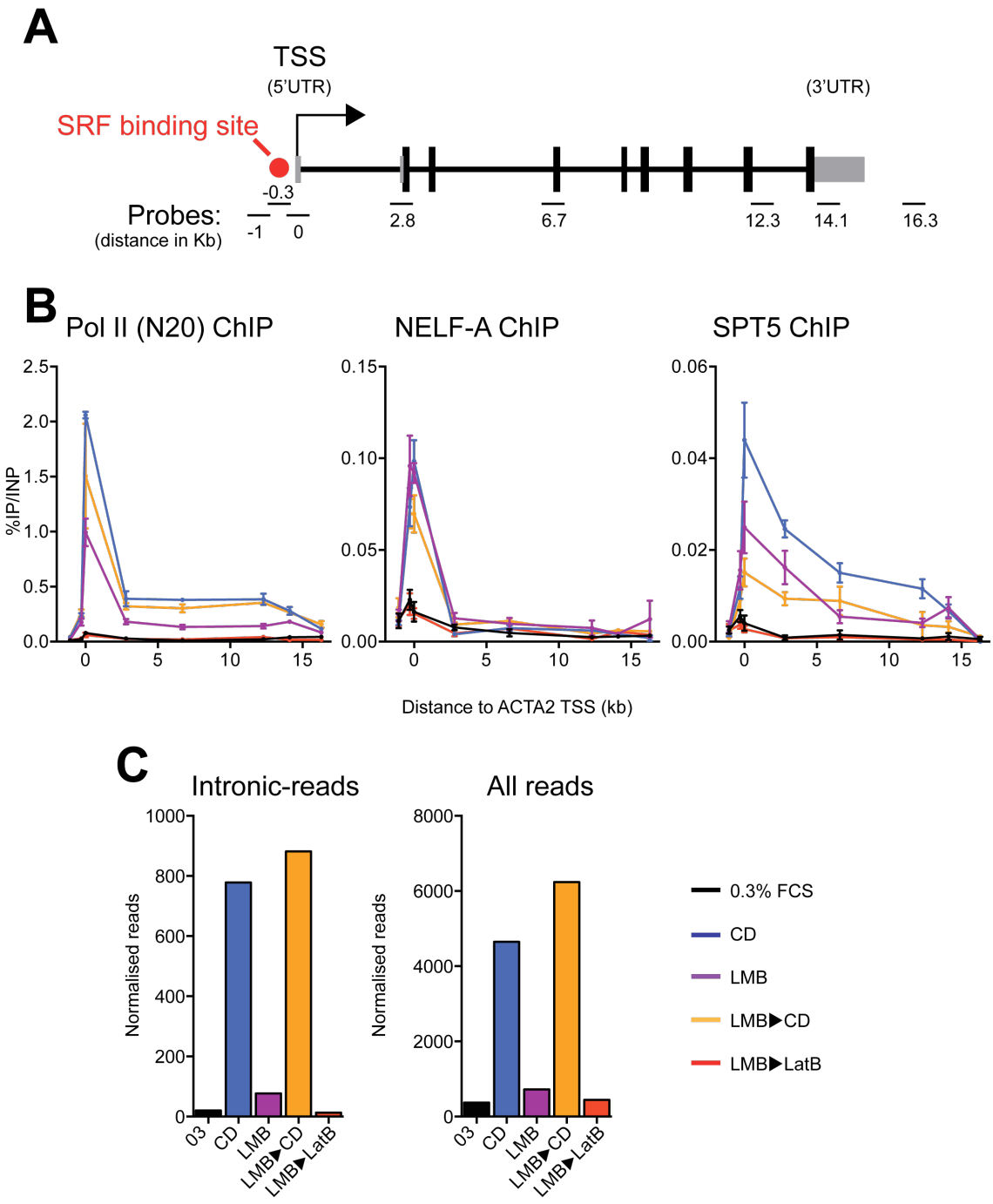


Figure 4.1 MRTF nuclear accumulation is sufficient for Pol II recruitment and escape at Acta2 gene.

(A) Representation of the *Acta2* gene illustrating the positions of different probes along the gene. Exons are indicated with 5'UTR and 3'UTR in grey and coding sequences in black; SRF binding site is indicated with a red circle. **(B)** ChIP of total Pol II, NELF-A and SPT5 at *Acta2* genes. **(C)** RNA-seq normalised read counts either total (right) or mapping in intronic features (left) of Acta2 gene.

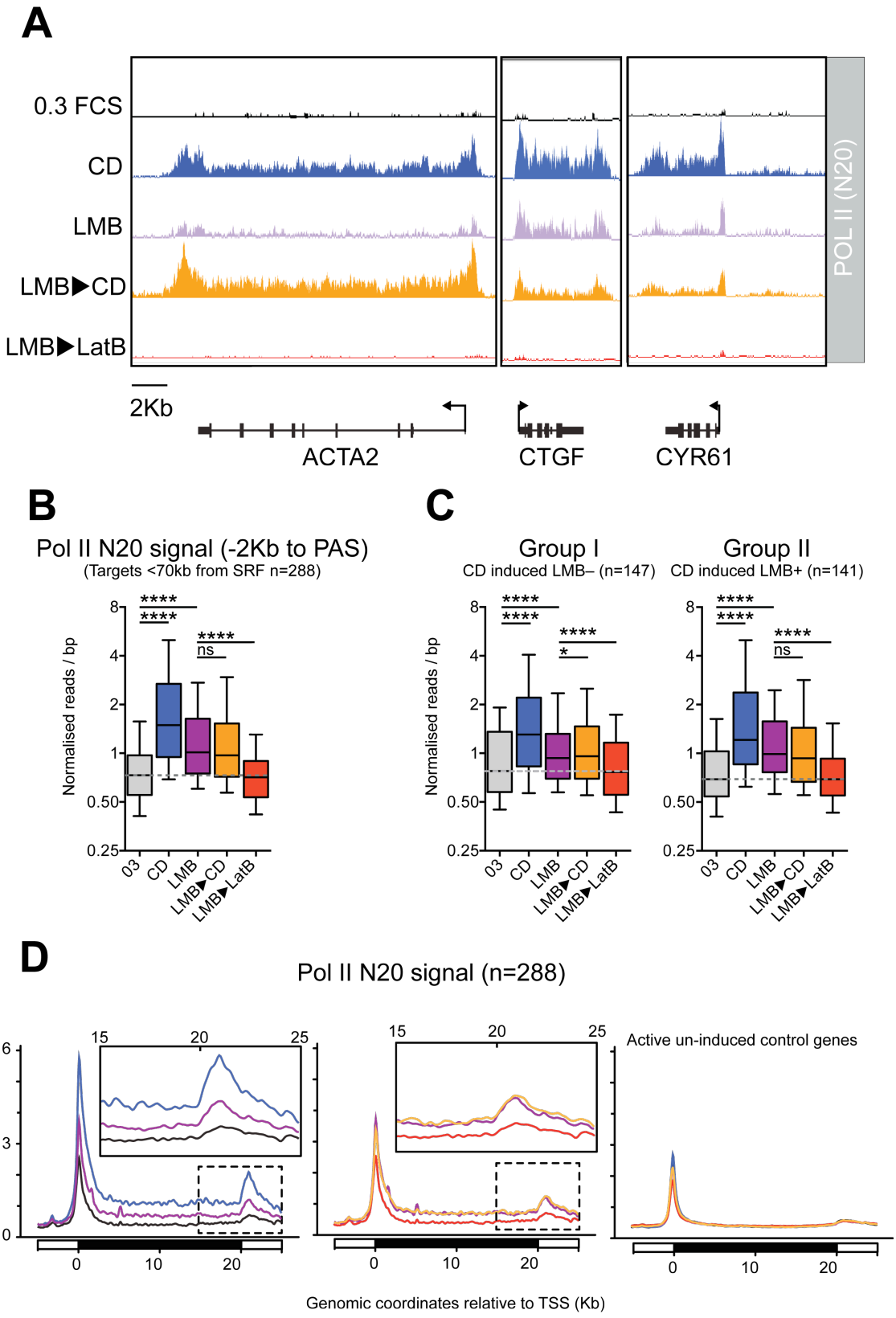

A
0.3 FCS
CD
LMB
LMB▶CD
LMB▶LatB
POL II (N20)
2Kb
ACTA2
CTGF
CYR61
B
Pol II N20 signal (-2Kb to PAS)
(Targets <70kb from SRF n=288)
Normalised reads / bp
C
Group I
CD induced LMB– (n=147)
Group II
CD induced LMB+ (n=141)
03
CD
LMB
LMB▶CD
LMB▶LatB
D
Pol II N20 signal (n=288)
Normalised count per base
Active un-induced control genes
Genomic coordinates relative to TSS (Kb)

Figure 4.2 MRTF nuclear accumulation is sufficient for Pol II recruitment and escape at SRF-MRTF target genes.

(A) Representative Pol II (N20 antibody) ChIP-seq tracks on *Acta2, Ctgf* and *Cyr61.* **(B)** Nuclear accumulation of MRTF shows induced recruitment and escape of Pol II at all 288 MRTF-target genes. Pol II ChIP-seq read counts from -2Kb to pA site per bp. **(C)** Comparison of Pol II recruitment and escape at group I-II genes, as described in Chapter 2. Group I are genes out of the 288 selected targets showing no LMB induction while group II are genes out of the 288 selected target showing partial induction after LMB stimulation. (B and C) The middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90^{th} and 10^{th} percentiles. Statistical significance, Wilcoxon test, (*) $P < 0.05$ (****) $P < 0.0001$, (ns) non-significant**. (D)** Metaprofile of Pol II ChIP-seq of the 288 MRTF-target genes selected from RNA-seq studies in Chapter 2. Normalised ChIP-seq read counts are shown across gene loci, standardised to 20kb, and flanking 5 kb. Dashed box represents the zoom in towards the end of the metaprofiles from 15Kb from the theoretical TSS down to +5 Kb. The third panel from the left represents constitutively active un-induced genes.

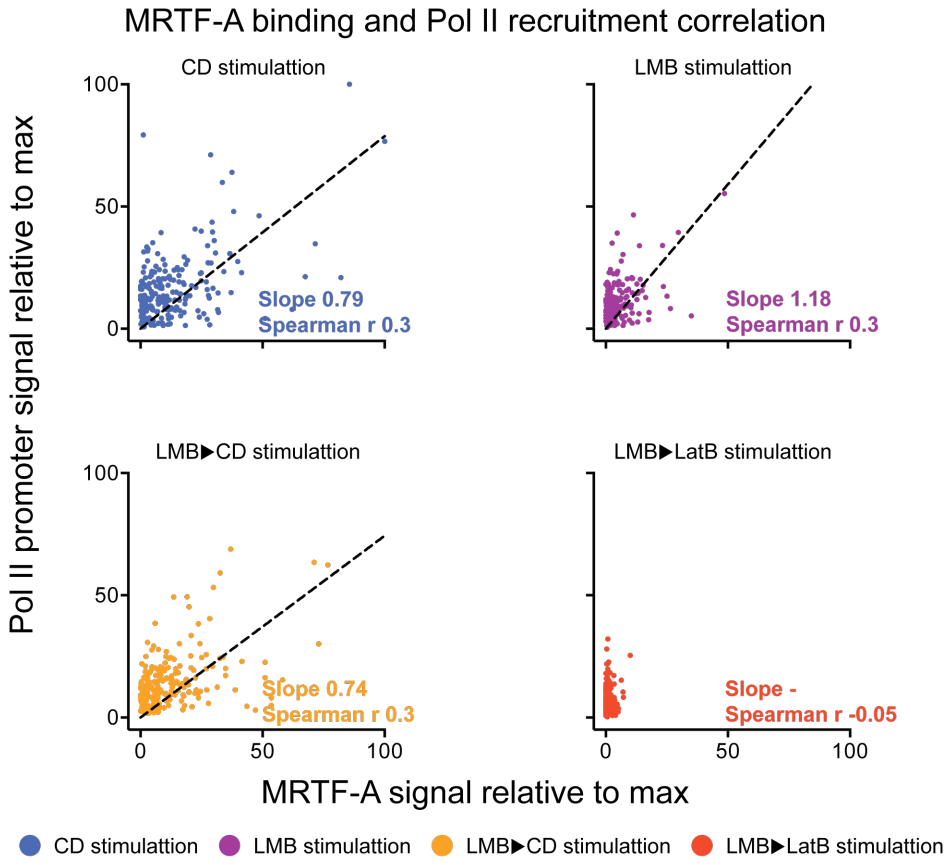


Figure 4.3 Correlation between MRTF binding and Pol II recruitment at target gene promoters.

MRTF-A binding intensity correlate with recruited Pol II at target gene promotes in CD (blue), LMB (purple), LMB▶CD (orange) but not in LMB▶LatB (red) conditions. The highest closest MRTF binding site in CD was selected per target gene. Both Pol II signal and MRTF-A were scaled to the highest signal observed across conditions. The significance of the Spearman r shows value with $P < 0.001$ for CD, LMB and LMB▶CD correlations. The LMB▶LatB did not show any significant correlation.

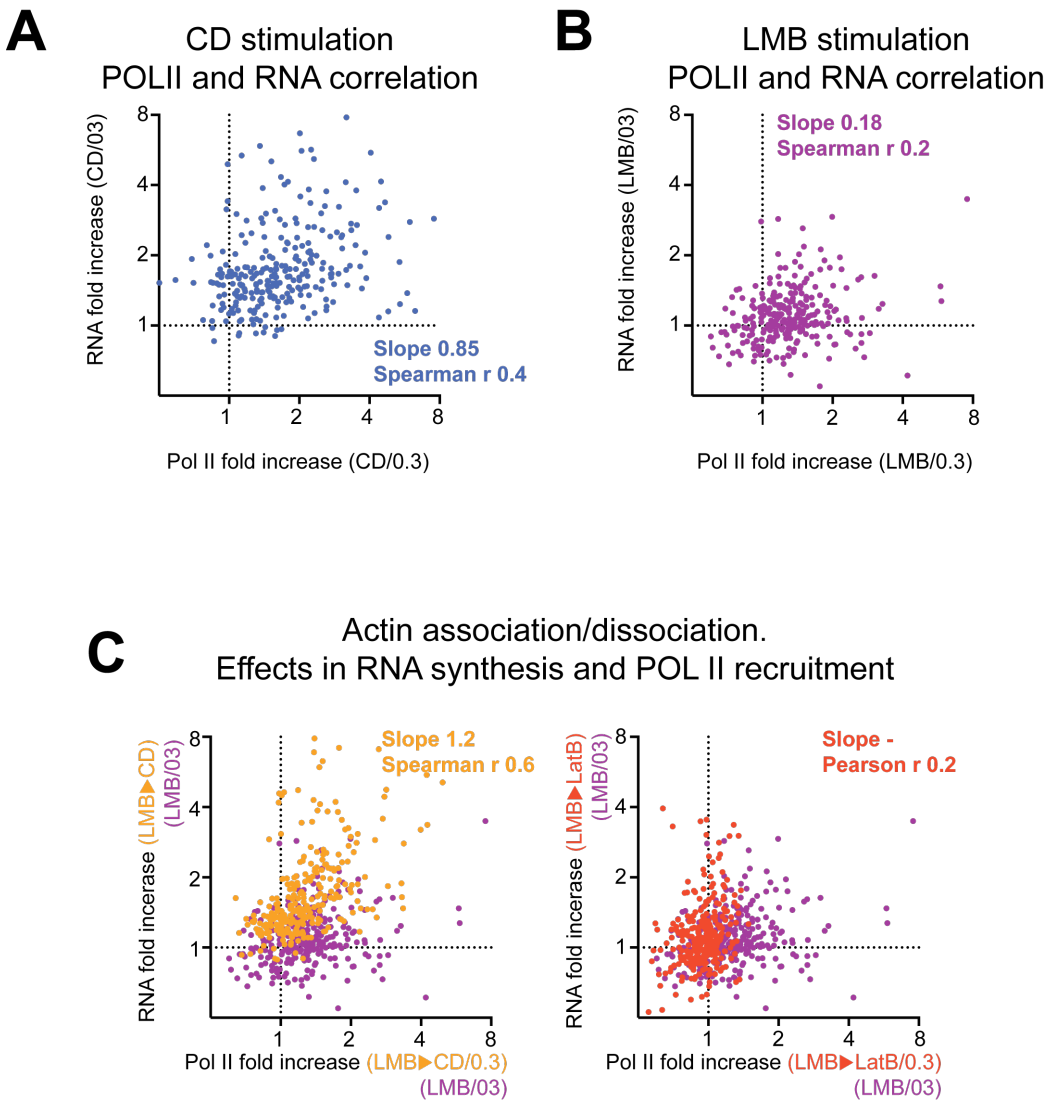


Figure 4.4 Correlation between travelling Pol II and RNA production.

(A and B) Correlation between increase in Pol II signal within the gene body of target genes and the fold induction observed at the RNA level. Data obtained from total Pol II ChIP-seq was crossed with the RNA-seq data described in Chapter 2. The fold increase was normalised to the highest increase observed across conditions. Graph A (blue) shows the correlation between Pol II and RNA induction after CD stimulation. Graph B (purple) shows the correlation between Pol II and RNA induction after LMB stimulation. **(C)** As in graph A and B shows the relation between Pol II and RNA increase when CD or LatB are added following the LMB stimulation.

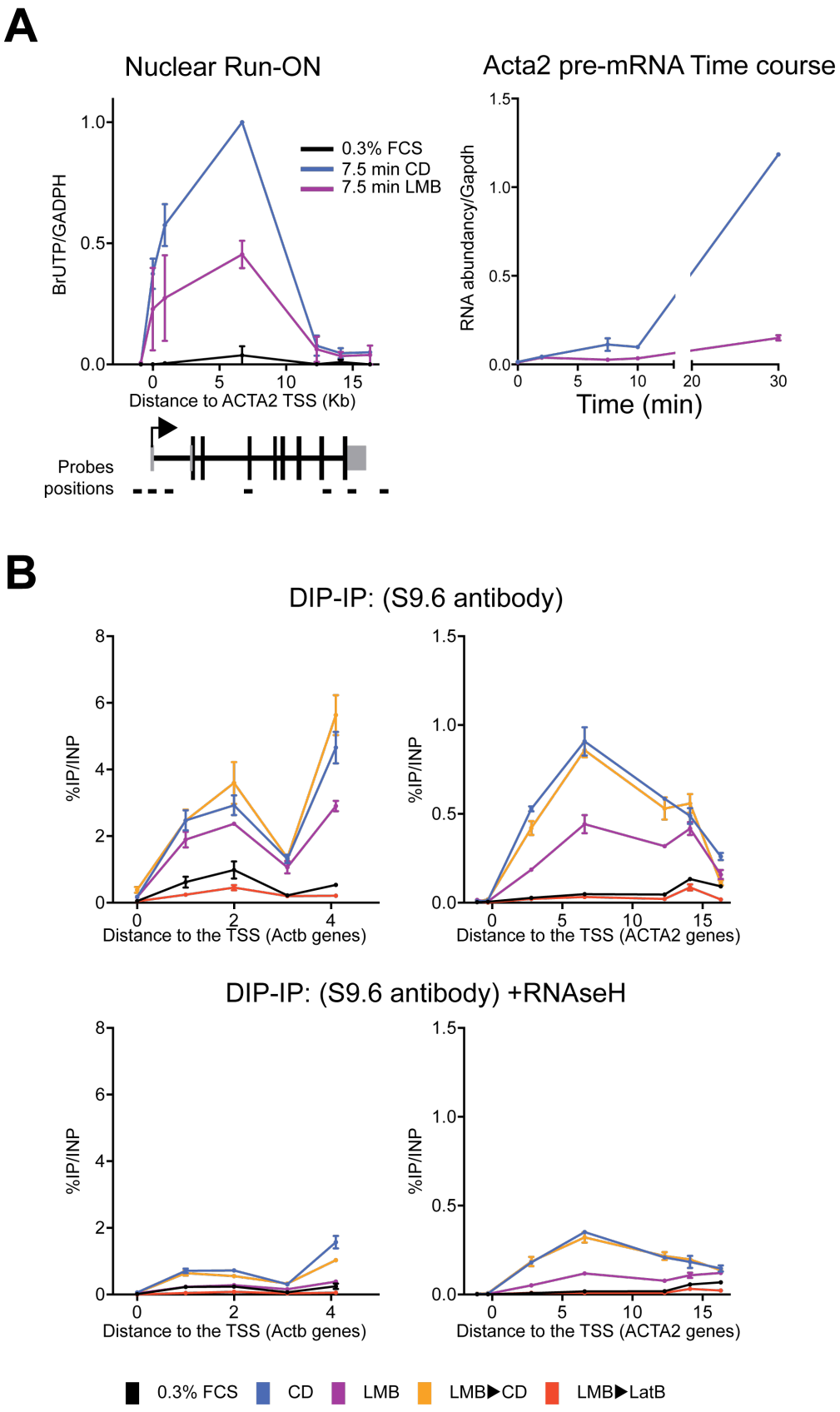
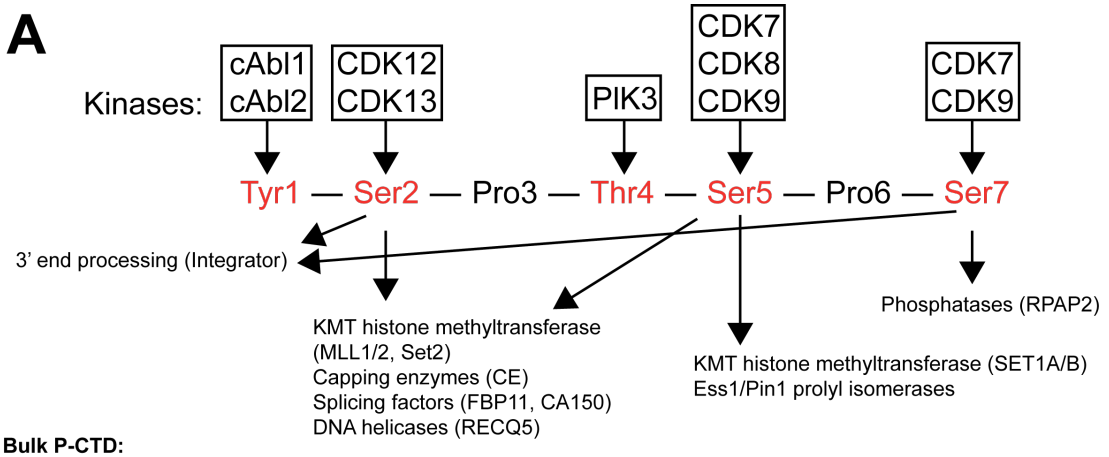

A
Nuclear Run-ON
BrUTP/GADPH
1.0
0.5
0.0
0.3% FCS
7.5 min CD
7.5 min LMB
0
5
10
15
Distance to ACTA2 TSS (Kb)
Probes positions
Acta2 pre-mRNA Time course
RNA abundancy/Gapdh
1.5
1.0
0.5
0.0
0
5
10
20
30
Time (min)
B
DIP-IP: (S9.6 antibody)
%IP/INP
8
6
4
2
0
0
2
4
Distance to the TSS (Actb genes)
1.5
1.0
0.5
0.0
0
5
10
15
Distance to the TSS (ACTA2 genes)
DIP-IP: (S9.6 antibody) +RNAseH
%IP/INP
8
6
4
2
0
0
2
4
Distance to the TSS (Actb genes)
1.5
1.0
0.5
0.0
0
5
10
15
Distance to the TSS (ACTA2 genes)
0.3% FCS
CD
LMB
LMB▶CD
LMB▶LatB

Figure 4.5 RNA synthesis following MRTF nuclear accumulation during productive and unproductive stimulation.

(A) (*Left*) nuclear run-on assay on the nascent *Acta2* transcript. Gene structure and probe positions are indicated at the bottom of the panel. Nuclei were extracted following 7.5 minutes of either CD (blue) or LMB (purple) stimulation (see materials and methods). (*Right*) Intronic *Acta2* RNA accumulation time course following either CD or LMB stimulation. Total RNA was extracted with conventional protocol and analysed by RT-qPCR. **(B)** DIP-IP on either *Actb* (left) or *Acta2* (right) genes using the S9.6 antibody. Bottom two panels show the DIP signal following RNase H treatment.



Bulk P-CTD:
Chromatin remodellers (Chd8); Chromatin chaperons (Spt6); Splicing factors (U2AF65, SCAF8, SR proteins); Cleavage and PA (CstF50, CPSF)

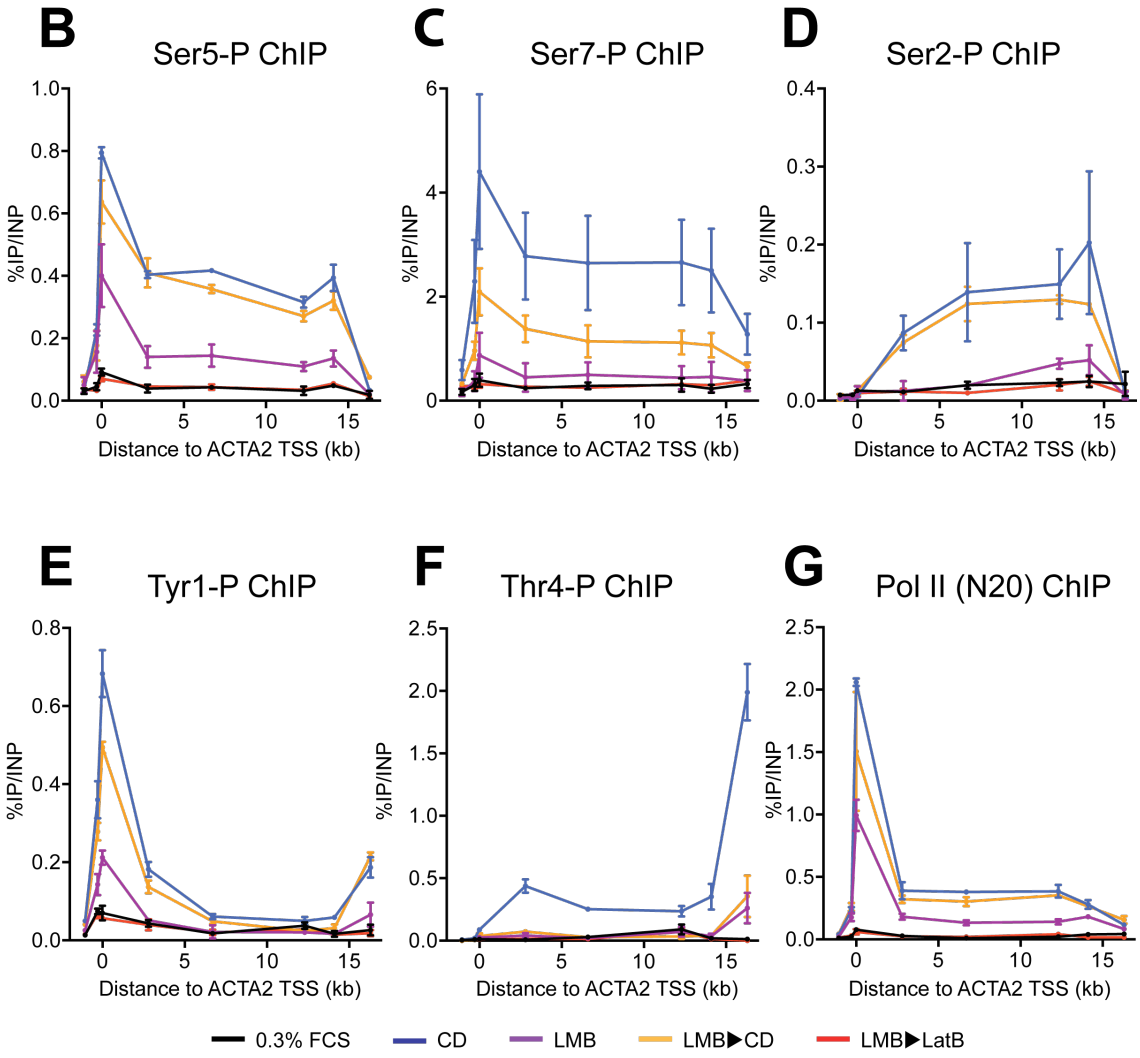


Figure 4.6 MRTF nuclear accumulation is not sufficient for correct Pol II phosphorylation at Acta2 gene.

(A) Schematic representation of the heptad sequence within Pol II C-terminal domain. In red are highlighted the residues that can be modified by phosphorylation. At the top within each box are listed the kinases known to modify the presented residues. At the bottom are listed the functions that have been reported to correlate with the modified residues. **(B to F)** ChIP experiments at *Acta2* model gene of different phospho-specific antibodies recognising defined residues: Serine 5-P (H14 antibody), Serine 7-P (4E12), Serine 2-P (H5), Tyrosine 1 (3D12), Threonine 4-P (6D7). **(G)** Pol II ChIP as in Figure 4.1 for comparison.

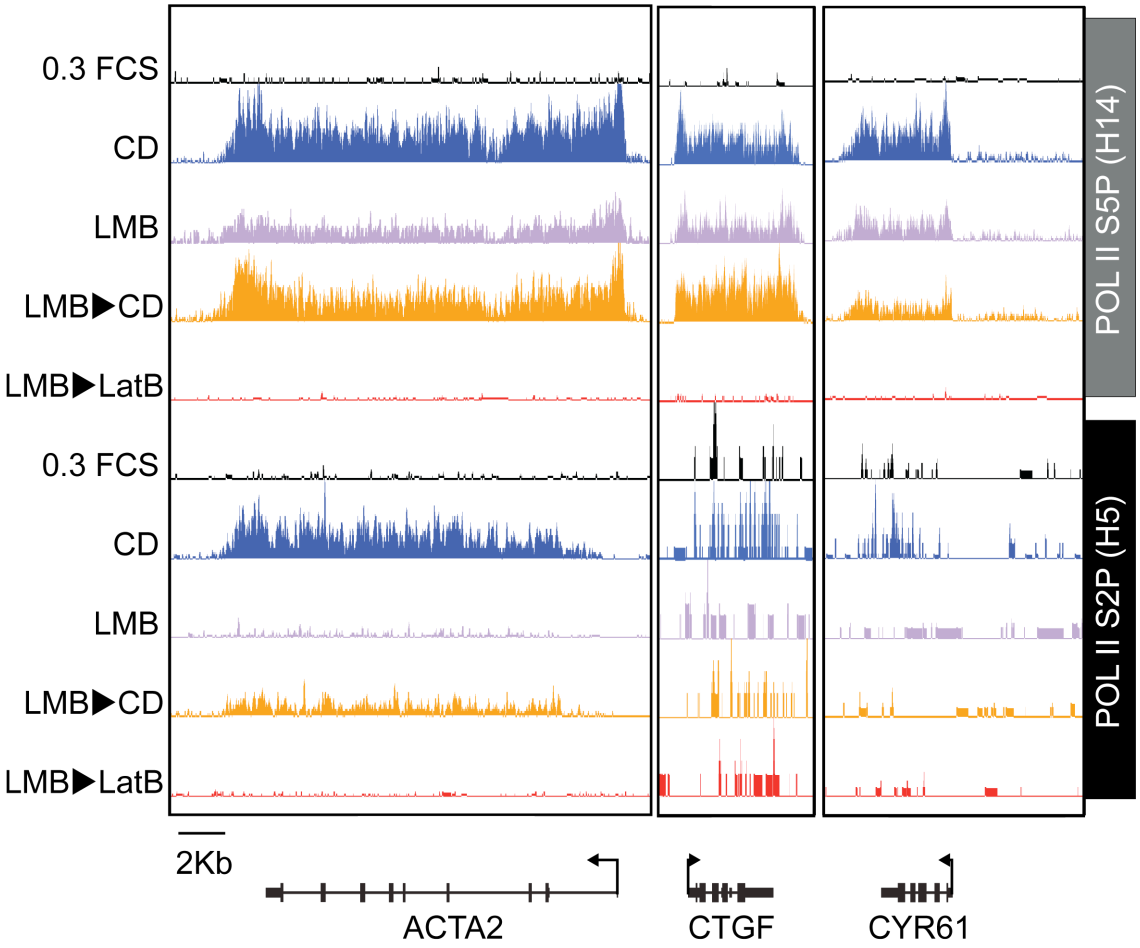


Figure 4.7 Serine 5 and Serine 2 ChIP-seq across conditions.

Representative Serine 5 phosphorylation (S5P) and Serine 2 phosphorylation (S2P) ChIP-seq on *Acta2, Ctgf* and *Cyr61* genes.

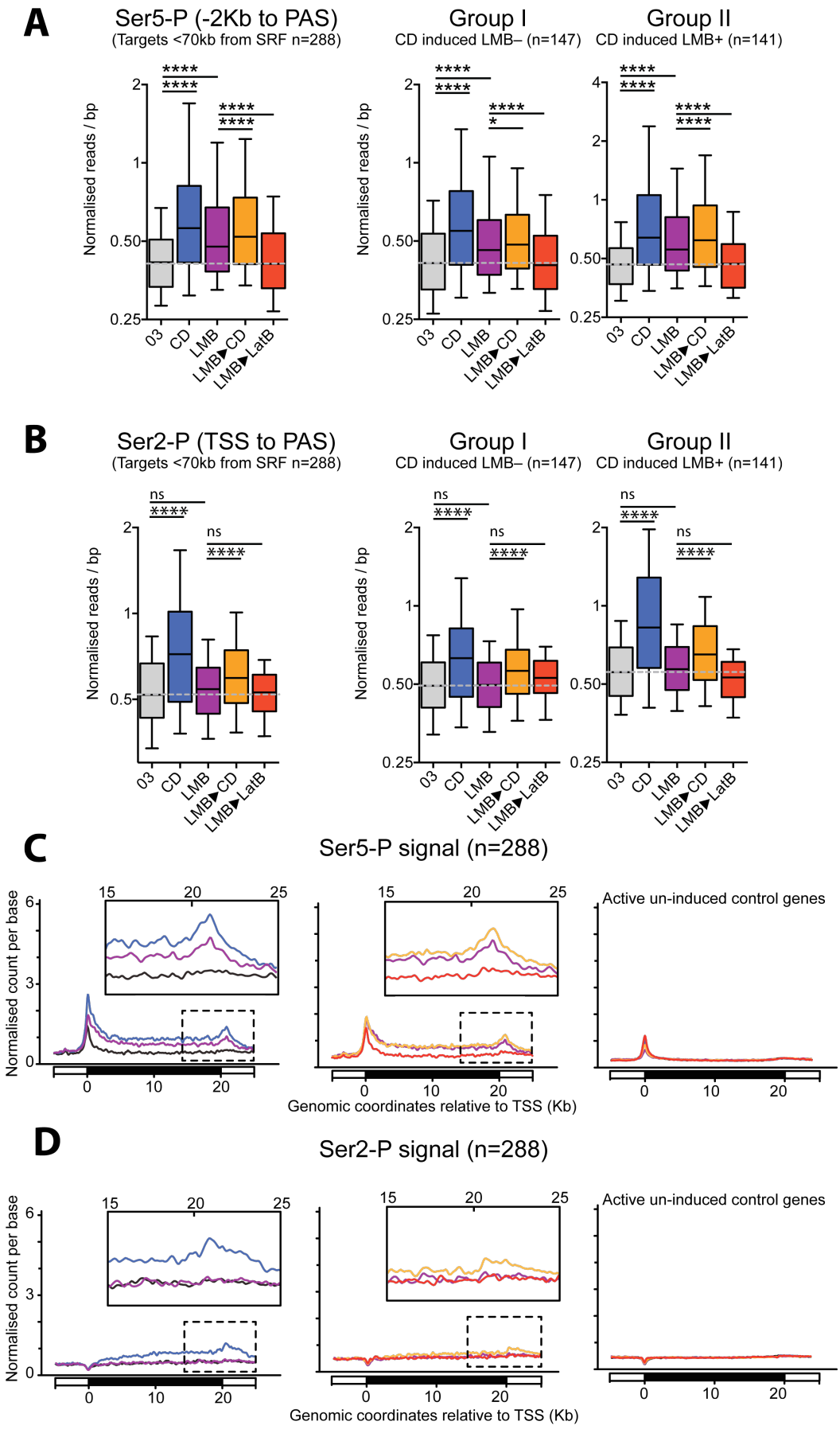

A
Ser5-P (-2Kb to PAS)
(Targets <70kb from SRF n=288)
Group I
CD induced LMB– (n=147)
Group II
CD induced LMB+ (n=141)
Normalised reads / bp
03
CD
LMB
LMB▶CD
LMB▶LatB
B
Ser2-P (TSS to PAS)
(Targets <70kb from SRF n=288)
Group I
CD induced LMB– (n=147)
Group II
CD induced LMB+ (n=141)
Normalised reads / bp
C
Ser5-P signal (n=288)
Normalised count per base
Active un-induced control genes
Genomic coordinates relative to TSS (Kb)
D
Ser2-P signal (n=288)
Normalised count per base
Active un-induced control genes
Genomic coordinates relative to TSS (Kb)

Figure 4.8 MRTF nuclear accumulation is sufficient for Ser5 but not Ser2 phosphorylation.

(A) Comparisons of the average ChIP densities of Serine 5-P (H14 antibody) from -2Kb to pA site for the 288 MRTF-SRF selected targets (left panel), or Group I and II as defined in Chapter 2 across conditions. **(B)** As panel A but assessing Serine 2-P (H5 antibody). **(A and B)** The middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90^{th} and 10^{th} percentiles. Statistical significance, Wilcoxon test, (*) $P < 0.05$ (****) $P < 0.0001$, (ns) non-significant**. (C)** Metaprofile of Pol II Serine 5-P ChIP-seq of the 288 MRTF-target genes selected from RNA-seq studies in Chapter 2. Normalised ChIP-seq read counts are shown across gene loci, standardised to 20kb, and flanking 5 kb. Dashed box represents the zoom-in towards the end of the metaprofiles from 15Kb from the theoretical TSS down to +5 Kb from the PAS. The third panel from the left represents constitutively active un-induced genes. **(D)** As in panel C but showing Pol II Serine 2-P CHIP-seq reads counts merged into a metaprofile.

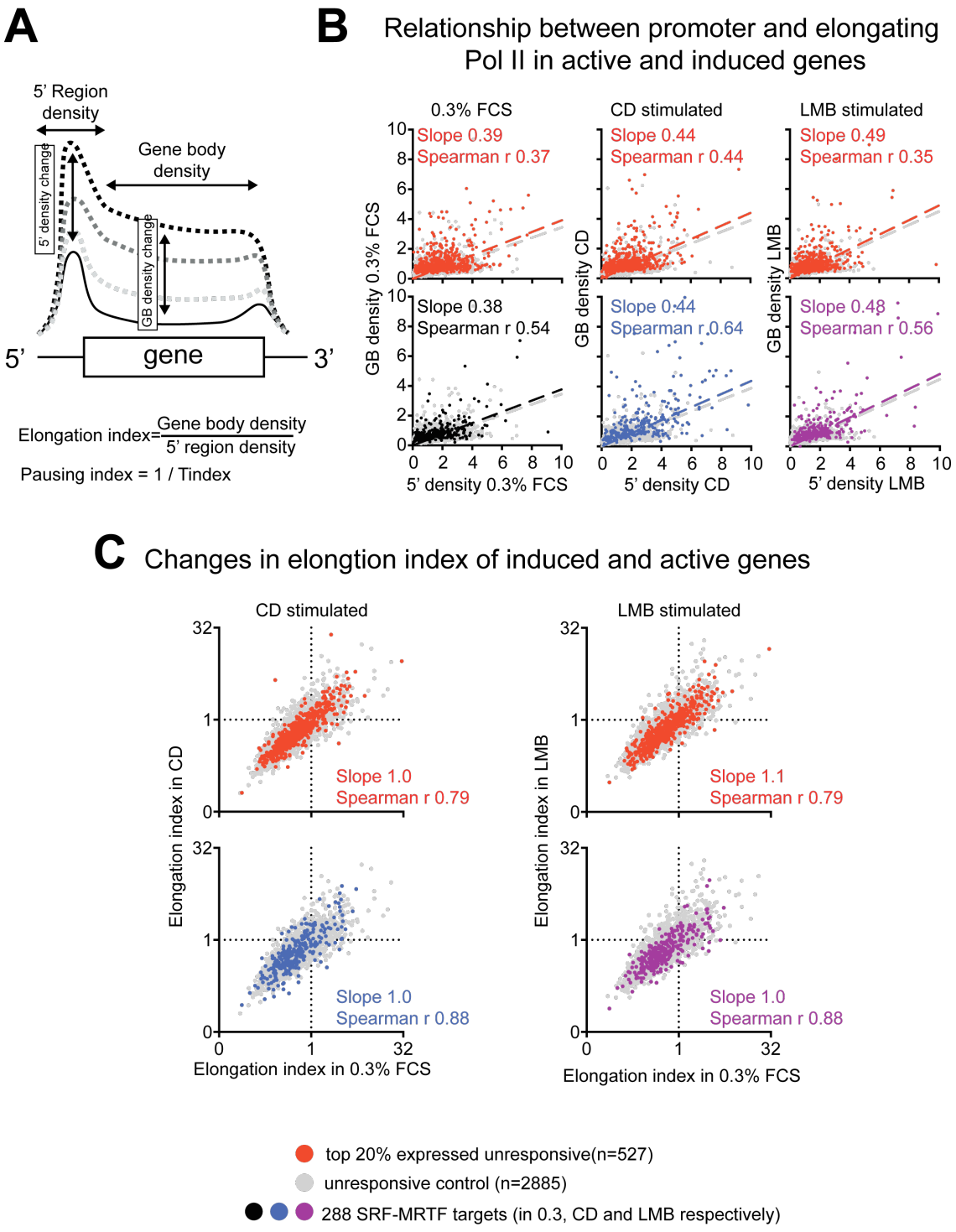

A
5' Region density
Gene body density
5' density change
GB density change
5'
gene
3'
Elongation index=Gene body density / 5' region density
Pausing index = 1 / Tindex
B
Relationship between promoter and elongating Pol II in active and induced genes
0.3% FCS
CD stimulated
LMB stimulated
Slope 0.39
Spearman r 0.37
Slope 0.44
Spearman r 0.44
Slope 0.49
Spearman r 0.35
Slope 0.38
Spearman r 0.54
Slope 0.44
Spearman r 0.64
Slope 0.48
Spearman r 0.56
GB density 0.3% FCS
GB density CD
GB density LMB
5' density 0.3% FCS
5' density CD
5' density LMB
C
Changes in elongtion index of induced and active genes
CD stimulated
LMB stimulated
Slope 1.0
Spearman r 0.79
Slope 1.1
Spearman r 0.79
Slope 1.0
Spearman r 0.88
Slope 1.0
Spearman r 0.88
Elongation index in CD
Elongation index in LMB
Elongation index in 0.3% FCS
Elongation index in 0.3% FCS
top 20% expressed unresponsive(n=527)
unresponsive control (n=2885)
288 SRF-MRTF targets (in 0.3, CD and LMB respectively)

Figure 4.9 Nuclear MRTF controls Pol II entry and escape.

(A) Schematic illustrations representing possible changes in Pol II profiles after stimulation. We considered changes in the 5' region density (-2Kb to 250bp) and gene body density (250bp to PAS) as defined by others (Min et al. 2011). Elongation rate is defined as the ratio of gene body density to pause peak density (that is the reverse of the Pausing index). **(B)** Relation between 5' density and gene body (GB) density per condition at (*top*) top 20% constitutive expressed in red (n=527) and (*bottom*) target genes (n=288) (black for 0.3% FCS, blue for CD and purple for LMB). In grey within each graph are all the constitutive active genes (n=2885). Dashed line represent the linear regression excluding outliers for each group of genes. **(C)** Correlation between elongation indexes in resting and stimulated condition for (*top*) top 20% constitutively expressed in red (n=527) and (*bottom*) target genes (n=288) (blue for CD and purple for LMB). In grey within each graph are all the constitutively active genes (n=2885).

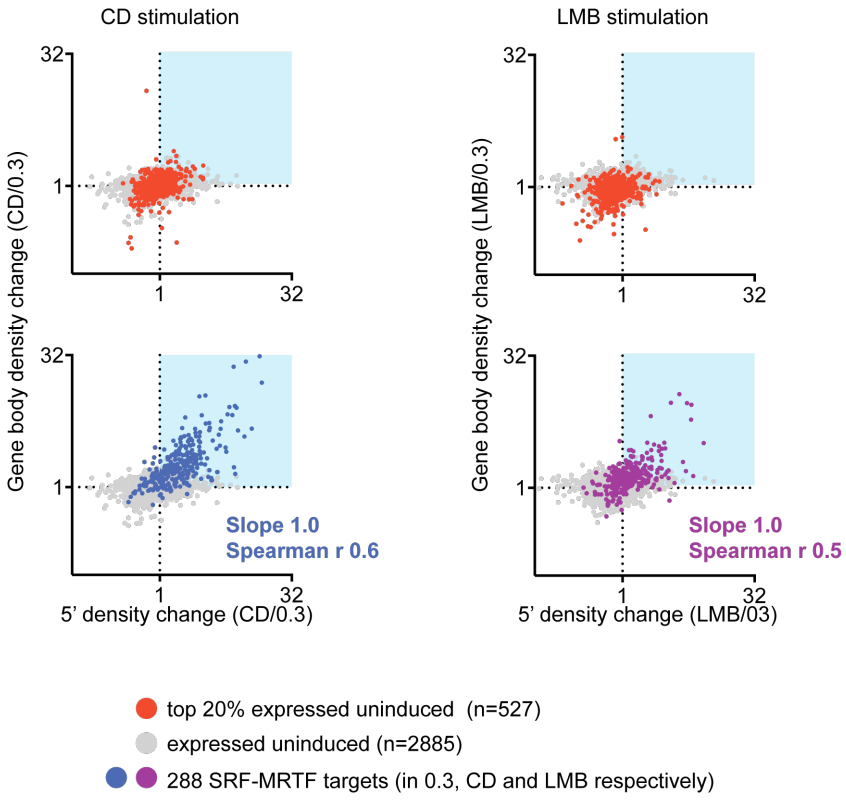


Figure 4.10 Nuclear MRTF equally enhances Pol II recruitment and escape.

Correlation between changes at the pause Pol II peak and Gene body Pol II signal (as CD or LMB stimulated over 0.3% FCS). (*Top*) in red top 20% constitutively expressed genes (n=527) and in grey all the constitutive genes (n=2885). (*Bottom*) coloured in blue (CD stimulation) or purple (LMB stimulation) are the 288 SRF/MRTF-targets while in grey are all the constitutively expressed genes (n=2885).

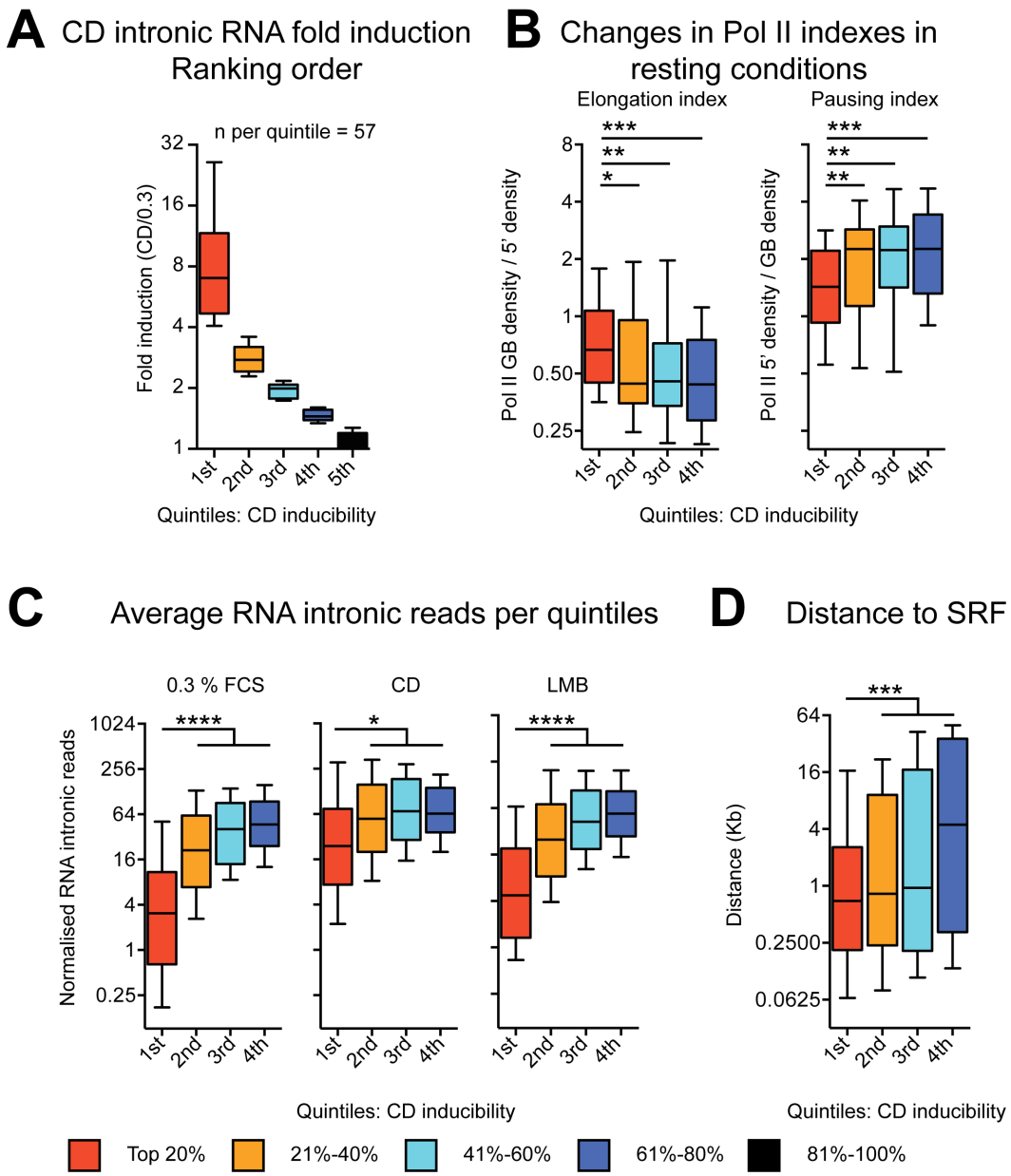


Figure 4.11 Paused Pol II inversely correlates with gene inducibility and SRF proximity

(A) RNA-seq fold induction in CD ranking per quintiles using intronic reads. **(B)** Elongation (*left*) and Pausing (*right*) indexes changes in 0.3% FCS for top 80% quintile as defined in panel A. **(C)** RNA intronic reads for top 80% quintiles in 0.3% FCS (*left*), CD (*centre*) and LMB (*right*). **(D)** SRF distance to genes' TSS per quintile. The middle line in each box plot indicates the median value, the top and bottom edges of the box plot are the 75th and 25th percentiles, and the small horizontal bars denote the 90th and 10th percentiles. (*) P < 0.05, (**) P< 0.01, (***) P < 0.001, (****) P < 0.0001, Mann-Whitney test.

Chapter 5. TCF-dependent chromatin signature in response to Ras activation

5.1 Aim

In recent years several chromatin changes have been reported to be a hallmark of transcriptional outcomes (Bernstein et al. 2007). Correlation studies based on genome-wide approaches have defined signatures that could potentially predict the transcriptional blueprint of a defined cell (J. Ernst et al. 2011). How these signatures are established is of great interest in order to understand the relationship between causes and effects and refine their predictive potential.

Several reports describe how signal influences chromatin changes, including histone modifications and nucleosome remodelling, often taking the role of TFs for granted (see Chapter 1). In this Chapter I am going to describe how the transcription factors TCF control the establishment of a defined chromatin signature in response to Ras-Erk signalling. The characterisation of the relationship between signal transduction, chromatin changes and the concomitant activation of defined target genes will be addressed.

Mouse Embryonic Fibroblasts (MEFs) lacking all three TCFs (named ESN standing for, Elk-1$^{-/-}$SAP-1$^{-/-}$Net$^{\partial/\partial}$) were previously generated in the lab (Costello et al. 2010) providing a clean tool to define targets induced by specific stimuli. In this chapter, having defined a *bona fide* set of targets and the stimuli that specifically activates the TCFs, I am going to describe how the TCFs control both chromatin changes and assembly of the transcriptional machinery. Finally, using reconstitution experiments, I am going to uncouple signal-dependent chromatin changes and mechanisms in transcriptional activation. The pivotal role of the Elk-1 activation domain will be emphasised.

This project was done in collaboration with Cyril Esnault.

5.2 TCFs but not MRTFs are required to induce IEGs in response to TPA

The Ternary complex factors (TCFs) are nuclear MAPK substrates. Each TCF protein is able to form a ternary complex with SRF at IE genes such as *c-Fos,*

Egr1, Egr2 and others (Dalton & Treisman 1992; Esnault et al. 2014). Although TCFs can act redundantly with other ETS proteins independently of SRF (Hollenhorst et al. 2011), a defined subset of SRF-dependent genes are regulated in response to MAPK activation and require TCFs' functions (Esnault et al. 2014). TCF-dependent gene activation involves the recruitment of subunits of the Mediator complex, Pol II entry and escape in response to Erk activation (G. Wang et al. 2005; Stevens et al. 2002). Erk activation has been shown to favour defined chromatin modifications and gene activation in other contexts (Vicent et al. 2006).

We used the RAS/Erk/TCF signalling pathway as a working model in order to dissect specifically the establishment of active chromatin signatures in response to ectopic stimulation. Mouse Embryonic Fibroblasts (MEF) available in the lab lacking Elk-1, SAP-1 and Net activities (named ESN standing for, Elk-1$^{-/-}$SAP-1$^{-/-}$ Net$^{\partial/\partial}$) provided us with a powerful tool to study the mechanism of transcriptional activation and chromatin modification. As described in Chapter 7 this cell line shows a reduced growth rate and a block in cell cycle. Reconstitution experiments with Elk-1 suggest that all three members are required for normal proliferation, as Elk-1 alone seems insufficient (see chapter 7 for full description and chapter 8 for discussion).

In order to define the adequate stimulus to specifically activate the TCFs, Cyril Esnault compared serum and TPA stimulation (12-O-Tetradecanoylphorbol-13-acetate, a potent PKC activator able to induce Ras stimulation; see Chapter 1) in wild type and ESN MEFs. A panel of eight genes were selected, including known SRF/TCF-dependent (*Egr1, Egr2, Fos, Egr3, Ier2* and *Nr4a1*) and SRF/MRTF-dependent genes (*Srf* and *Vcl*). TPA specifically activates the SRF/TCF-dependent gene set in wild type MEF leaving *Srf* and *Vcl* unaffected (Figure 5.1A and 5.2A). On the other hand ESN MEF showed no accumulation of RNA for both SRF/TCF and SRF/MRTF specific targets upon TPA stimulation (Figure 5.1A and 5.2A). We noticed that LatB and U0126 (a MEK specific inhibitor) had context specific effects. Genes such as *Egr1*, *Ier2*, Fos and Nr4a1 showed sensitivity to U0126 while no effect following LatB was observed. On the other hand genes like *Egr2* and *Egr3* were shown to be impaired by both compounds following TPA activation. Although this observation would imply that MRTF could also be involved in the activation of these genes, it is important to consider that in ESN MEF following TPA both *Egr2* and *Egr3* did not show efficient induction. By contrast serum stimulation induced

most of the SRF/TCF and SRF/MRTF-specific genes, in both wild type and ESN MEFs (Figure 5.1B and 5.2B). The serum response detected in ESN MEF at SRF/TCF targets showed LatB sensitivity, suggesting that the TCFs could be substituted with the MRTFs in response to Rho activation. This is supported by the fact that in ESN MEF the response to serum was not affected by U0126 treatments while a clear LatB sensitivity was recorded across all targets (Figure 5.1B). We therefore considered TPA a good way to specifically activate TCF-dependent target genes. As shown in Figure 5.2A TPA quickly activates IE genes allowing accumulation of mature mRNA within 45 minutes of stimulation.

In order to confirm that the defect observed in ESN MEF was uniquely due to a loss in transcriptional-related mechanism we compared the kinetic of Erk activation in wild type and ESN MEFs using a phospho-specific antibody for Erk1/2 (Figure 5.3A). ESN MEF, like the wild type cell line, showed a quick Erk1/2 activation within 5 minutes of TPA stimulation (Figure 5.3A and B). Furthermore, while in wild type MEF the quick activation was followed by a rapid shutdown, in ESN MEF the attenuation of active Erk1/2 was shown to be slower (Figure 5.3B). This phenomenon might be caused by loss of expression of Dusp phosphatases in ESN MEF. Dusp phosphatases are SRF-TCF targets involved in the down-regulation of MAP Kinases by de-phosphorylation (Kondoh & Nishida 2007).

The defect in activation could also be caused by a dishomogeneous response by the ESN MEF cell pool. To exclude this scenario we monitored the accumulation of *Egr1* protein using FACS analysis across the cell population. Stimulation with serum for one hour allowed accumulation of *Egr1* protein in both wild type and ESN MEFs confirming that the knock out cell line was transcriptionally competent and the response homogenous within the cell population (Figure 5.4A). On the other hand *Egr1* protein was not expressed in ESN MEF following TPA stimulation corroborating the already described transcriptional defect (Figure 5.4B).

5.3 Defined chromatin changes occur at *Egr1* promoter following activation

Activation of transcription correlates with defined chromatin changes including: histone H3 serine 10 phosphorylation (H3S10-P), a modification involved

in both transcription and cell division (S.-H. Yang et al. 2013); histone H3 lysine 4 trimethylation (H3K4me3), a modification associated with promoters (Guenther et al. 2007); histone H3 lysine 9 and 14 acetylation (H3K9K14ac), H3K27ac and H4K16ac, associated with active genes and enhancers (Creyghton et al. 2010; Taylor et al. 2013), and macroscopic changes in nucleosome density (Gilchrist et al. 2010).

In order to establish how chromatin changes occur in response to Ras activation we performed a chromatin IP time course of the listed chromatin modifications at the *Egr1* model gene (Figure 5.5A). Following activation a clear reduction in H3 signal was observed starting at 15 minutes and continuing at 30 minutes after TPA stimulation (Figure 5.5B). This change occurred mainly at the *Egr1* TSS and within its ORF, probably due to PIC assembly and Pol II release. H3S10-P was shown to be the fastest mark to occur at the SRF-TCF binding site of *Egr1* promoter (Figure 5.5C). Most of the H3S10-P signal was already detected 5 minutes after TPA stimulation. Due to the possibility of epitope masking caused by acetylation of K9, further analyses are required. The use of different H3S10-P antibodies and antibodies able to recognise H3 N-terminal tail modified at both S10 and K9 will be shown later in this chapter.

Acetylation of lysine residues in positions 27, 9 and 14 of the histone H3 and lysine 16 of histone 4 showed similar kinetics with most signal detected 15 minutes after TPA stimulation and maintained at 30 minutes (Figure 5.5D, E and F). H3K27ac and H4K16ac occurred mainly at the SRF-TCF binding site while H3K9K14ac signal spread down to the TSS (Figure 5.5F). H3K4me3 was shown to be the most stable mark, being detectable also in resting conditions (Figure 5.5G). Noticeable changes occurred 30 minutes after TPA stimulation where the enhanced signal, like H3K9K14Ac, spanned from the SRF-TCF binding site down to the *Egr1* TSS.

Using this time course we also assessed Mediator and Pol II recruitment kinetics (Figure 5.6A). As showed in Figure 5.6 Med1, a major subunit of the Mediator complex, was recruited 15 minutes after TPA stimulation while Pol II was observed at 30 minutes (Figure 5.6A and B). The recruitment of Pol II at 30 minutes correlated with *Egr1* RNA precursor accumulation, showing a marked spike at 30 minutes following TPA stimulation (Figure 5.6C).

Taken together these observations show that TPA stimulation induces a set of chromatin modifications at *Egr1* promoter. These modifications occur with a defined kinetic and spatial distribution. Mediator recruitment occurs concomitantly with the establishment of the chromatin signature. Finally Pol II recruitment and release occur in a separate time frame than the assessed histone modifications and correlates with the accumulation of RNA precursor.

5.4 Signal induced chromatin changes at *Egr1* promoter require the TCFs

Having defined a set of chromatin modifications induced upon TPA stimulation we asked whether this chromatin signature required the presence of the TCFs. Using the *bona fide Egr1* gene we assessed the distribution and appearance of the defined set of chromatin modifications comparing wild type and ESN MEFs. We considered 0.3% FCS and 30 minutes following TPA as informative time points. As described before, 30 minutes is the time point where we could observe changes at each selected chromatin mark, Mediator recruitment, Pol II recruitment and escape and importantly, accumulation of pre-mRNA (see previous section).

As described, *Egr1* expression was initially shown to be defective in ESN MEF when compared to the wild type cell line (Figure 5.7A). This defect was observed with both pre-mRNA and mature mRNA. Furthermore, not only were the ESN cells unable to activate *Egr1* in response to TPA, they also showed a reduced baseline in resting condition. This defect was consistent with a loss in Pol II recruitment and escape, while SRF binding was mostly unaffected (Figure 5.7B). Occupancy of histone H3 in wild type and ESN MEFs was significantly different (Figure 5.7C). In ESN MEF overall H3 showed a higher signal than in wild type cells and upon TPA stimulation there was no change in H3 density within the ORF, corroborating the missing Pol II. In response to TPA H3S10-P, H3K27ac, H4K16ac and H3K9K14ac showed no change in ESN MEF while, as described before, in a wild type context a marked increased was measured at the SRF/TCF binding site with H3K9K14ac spreading down to the TSS (Figure 5.7D to G). H3K4me3 as previously described was the least changing chromatin mark in response to TPA. In ESN MEF it was possible to see an overall reduction of H3K4me3 (Figure 5.7H).

In conclusion TCFs are essential to activate *bona fide* targets and establish a signal-dependent chromatin signature at the *Egr1* promoter. In the following sections I am going to provide further insight using reconstitution experiments of the ESN MEF using Elk-1 wild type and different mutants.

5.5 IE gene activation requires two defined features in Elk-1 activation domain

TCFs are transcription factors equipped with a well-characterised activation domain (AD) towards their C-terminus. The role of the activation domain in transcription has been extensively studied. As reported by our lab and the Arnold Berk lab, a series of Ser/Thr sites decorate Elk-1 AD and are phosphorylated upon Ras-dependent Erk activation (Marais et al. 1993; Balamotis et al. 2009). These phospho residues are not the only elements required to induce transcription. Two key hydrophobic residues, embedded within the AD, are also required in order to activate target genes in response to signal. Studies conducted in the Berk lab elucidated the interplay between phospho and hydrophobic elements in coordinating the interaction with subunits of the Mediator complex (Balamotis et al. 2009). In particular Elk-1, a member of the TCF family, was shown to specifically interact with subunit 23 of the Mediator complex. Such an interaction occurs in a signal-dependent manner and requires both phospho and hydrophobic residues. Intriguingly, mutations or deletion of the hydrophobic residues (FW), although detrimental for Med23 interaction and gene activation, do not affect Elk-1 phosphorylation (Figure 5.8A).

In order to investigate the role of the Elk-1 activation domain in TCF-dependent gene activation I ectopically expressed Elk-1 variants in ESN MEF. As previously reported in the lab SAP-1 and Elk-1, but not Net, are functionally equivalent and can be substituted for thymocyte positive selection (Costello et al. 2010). ESN MEFs were infected with retroviruses expressing Elk-1 wild type, Elk-1 FW (variant with FW residues in position 378 and 379 deleted) or Elk-1 Nona (variant with alanine substitutions for each of the nine S/T-P site within Elk-1 AD), together with GFP (see Matherials and Methods). Each infected cell population was FACS-sorted into three subpopulations (low, medium and high) according to the GFP expression and cell pools expressing comparable amounts of Elk-1 protein

were selected for further analysis (Figure 5.8A and B). The knock out cell line was also infected with retrovirus harbouring only the vehicle as control.

Stimulation with TPA for 30 minutes reduced the mobility of the wild type and FW Elk-1 protein but left the Nona mutant unaffected (Figure 5.8B). This observation is consistent with a complete loss of Erk-dependent phospho-residues within the Elk-1 Nona AD. To further characterise the reconstituted cell lines I compared the binding of exogenous Elk-1 proteins expressed (Figure 5.9A). Chromatin IP experiments using a FLAG antibody did not show a detectable signal in any of the reconstituted cell lines (data not shown). The FLAG epitope was probably masked as fused to the N-terminal region of each mutant, close to the DNA binding domain. I therefore purified an in-house anti-mouse Elk-1 (aa309–429) antibody, able to recognise sequences at the C-terminal part of each mutant, against a peptide mapping across Elk-1 AD with alanine substitutions at each S/T-P site. As shown in Figure 5.9 A ChIP experiment using this antibody showed a comparable signal for each mutant at the *Egr1* promoter. Furthermore SRF binding was partially induced by each Elk-1 variant, if compared to the ESN MEF infected with vehicle only (Figure 5.9B). Furthermore TPA stimulation allowed Med1 and Pol II recruitment only in cells expressing Elk-1 wild type, consistent with the model that both phospho and hydrophobic residues are required for Mediator recruitment and target gene activation (Figure 5.10A and B).

Finally I characterised the transcriptional outcome of each reconstituted cell line compared to the parental wild type and ESN MEFs (Figure 5.11). Reconstitution with Elk-1 wild type allowed similar TPA-induced expression of *Egr1*, *Fos*, *Ier2* and *Egr2* if compared to parental MEF carrying all three TCF (Figure 5.11). On the other hand Elk-1 FW and Nona reconstituted cell lines did not show, in response to TPA, efficient gene induction (Figure 5.11).

In summary we established a set of reconstituted cell lines expressing comparable amounts of Elk-1 wild type, Elk-1 FW and Elk-1 Nona. Each Elk-1 mutant showed comparable binding efficiency and as previously reported only Elk-1 wild type was able to induce transcription via Mediator and Pol II recruitment. Mutation of the FW hydrophobic residues, although affecting Mediator recruitment and activation of transcription, allows Erk-dependent phosphorylation of Elk-1 AD.

5.6 Elk-1 wild type is sufficient to re-establish signal induced chromatin changes at *Egr1* promoter

The experiments described in section 5.3 strongly suggest that any signal-induced chromatin modification can occur only in the presence of defined TFs. Within this context the TCFs might work as anchoring factors, dictating the exact positioning of modifications in relation to the targeted gene.

Having established a set of reconstituted ESN MEF cell lines (see section 5.4) we initially assessed if Elk-1 wild type was sufficient to allow a correct chromatin signature at the *Egr1* promoter. Strikingly, wild type Elk-1 (Elk-1 WT) established a correct chromatin signature, comparable to wild type MEF (Figure 5.12). In particular H3 showed a reduced signal when the Elk-1 WT cell line was induced via TPA at *Egr1* TSS and within its ORF (Figure 5.12 B). H3S10-P, and co-occurrence of S10-P and K9Ac, showed a marked induction at the SRF/TCF binding site (Figure 5.12C and D). Similarly acetylation of residues K27, K9 and K14 of histone H3 together with K16 of histone H4 were induced upon TPA stimulation in ESN MEF rescued by Elk-1 WT (Figure 5.12 E to G). A similar H3K9K14ac profile was also achieved (Figure 5.12G). Lastly H3K4me3 was enhanced in ESN MEF expressing Elk-1 WT (Figure 5.12H).

In conclusion chromatin changes at the *Egr1* promoter relies on the TCFs. Elk-1 is sufficient to re-establish a correct response in terms of transcriptional activation and chromatin signature modifications. This framework provides us with an immense tool to dissect the relationship between signal, chromatin changes and transcription.

5.7 Contribution of Elk-1 features in establishing chromatin signatures at *Egr1* promoter

As introduced earlier the mechanism of transcriptional activation guided by the TCFs involves the recruitment of subunits of the Mediator complex that in turn allow Pol II promoter entry. This mechanism requires defined features within the TCF AD including S/T-P sites and hydrophobic residues. Elk-1 is the best-characterised TCF. Two defined residues (FW) surrounded by a series of S/T-P sites could be specifically targeted to impair Elk-1 mediated transcription without

affecting their association to genomic loci (see Chapter 1 and previous sections). These observations imply that the Ras induced signalling cascade can still reach the Elk-1 FW mutant at target gene promoters.

To our surprise it was possible to observe, upon TPA stimulation, the induction of several chromatin marks in cells reconstituted with Elk-1 FW (Figure 5.13). No change was observed at the level of histone H3 signal for both Elk-1 FW and Nona (Figure 5.13A). On the other hand a marked induction of H3S10-P, H4K16ac and H3K9K14ac was detected only at the SRF-TCF binding site in Elk-1 FW (Figure 5.13B,C, E and F). Elk-1 Nona showed no enhancement of these modifications. H3K27ac was overall higher in resting and stimulated conditions in both Elk-1 FW and Nona (Figure 5.13 D). Despite this higher baseline H3K27ac did not show any induction upon TPA in cells harbouring Elk-1 FW or Nona. In addition, direct comparison of H3K27ac signals highlights an overall impairment in cells reconstituted with the FW and Nona mutants (Figure 5.14 A). H3K4me3 was highly enhanced in both Elk-1 FW and Nona with a different distribution than with Elk-1 WT (figure 5.13 G and 5.14 A for comparison). Intriguingly H3K9K14ac also showed a different distribution in Elk-1 FW when compared to the Elk-1 WT (Figure 5.13 F and 5.14 A). In particular the TPA-dependent enhancement of this modification in Elk-1 FW was restricted at the SRF-TCF binding site while in the wild type context this modification spread towards *Egr1* TSS (figure 5.14 A).

Recently provocative publications proposed that defined chromatin signatures could per-se induce transcription via recruitment of the pTEF-b complex (Zippo et al. 2009). To this end we assessed the recruitment of CDK9 at the *Egr1* promoter following TPA stimulation in reconstituted ESN MEF. While cells expressing Elk-1 wild type showed detectable signal for p-TEFb at *Egr1* TSS, no signal could be detected in cells expressing Elk-1 FW or Nona (Figure 5.14 B).

In conclusion it is possible to uncouple defined chromatin changes from the transcription process. Chromatin modification can occur in response to Elk-1 DNA binding or as a result of Elk-1 AD phosphorylation at the SRF/TCF binding site. A complete shaping of the chromatin changes is achieved through recruitment of the transcription machinery that seems to enhance H3K4me3 and H3K9K14ac at *Egr1* TSS. The histone modifications observed in cells expressing Elk-1 FW are not sufficient to induce recruitment of p-TEFb at *Egr1* promoter.

5.8 Elk-1 activation domain is required to maintain a permissive chromatin at *Egr1* ORF

It has been recently reported that Erk2 and Elk-1 genomic distribution account for diverse transcriptional states in human ESc (Göke et al. 2013). Elk-1 and Erk2 co-localisation describe actively transcribed genes involved in cell cycle and proliferation while Elk-1 alone seems to enhance H3K27me3 and repress transcription. Lysine 27 tri-methylation at histone H3 (H3K27me3) is a modification associated with silenced and bivalent polycomb-associated chromatin (Bernstein et al. 2006). In order to assess if TCF activities provide an antagonistic mechanism to silenced chromatin we analysed the distribution of H3K27me3 in different cell contexts.

The signal for H3K27me3 was overall low for both ESN MEF and wild type MEF (Figure 5.15 A). H3K27me3 showed no enrichment at *Egr1* promoter and gene body in any of the conditions examined. This observation was consistent with the ability to stimulate IE genes in ESN MEF with serum instead of TPA, probably through MRTF activity (see section 5.1). To our surprise MEF expressing Elk-1 Nona showed the opposite, a striking enhancement of H3K27me3 that from the TSS was invading the *Egr1* gene body (Figure 5.14 B). ESN MEF infected with Elk-1 WT or FW did not show this enrichment in H3K27me3 suggesting that the S/T-P sites per-se were sufficient to maintain a permissive chromatin even in resting conditions. Elk-1 Nona possibly works as a dominant negative overcoming the positive effects of other TFs at the *Egr1* promoter. Recent publications have described how Erk constantly shuttles from the cytoplasm to the nucleus maintaining a basal level of transcription (Aoki et al. 2013). These stochastic events might be responsible for maintaining permissive chromatin at IE genes. Further studies are required to highlight the mechanisms responsible for H3K27me3 appearance in MEF expressive Elk-1 Nona and the possible role of polycomb in establishing this chromatin signature.

5.9 Summary

In this chapter I described how a defined chromatin signature could be established in response to TPA at the IE gene *Egr1*. All TCFs are required to

induce several changes at the *Egr1* promoter and reconstitution of knock out cells with Elk-1 wild type protein is sufficient to re-establish a correct chromatin signature and gene expression if compared to wild type MEF. The contributions of defined features embedded within Elk-1 AD were dissected highlighting the requirement of both S/T-P sites and the two hydrophobic residues FW for IE gene stimulation. Furthermore it was possible to disentangle chromatin changes and transcriptional activation. Phosphorylation of Elk-1 is sufficient per-se to enhance several chromatin modifications while others only require Elk-1 DNA binding activity. Furthermore it was possible to observe that in resting conditions intact S/T-P sites within the Elk-1 AD are required to maintain a permissive chromatin at *Egr1* promoter.

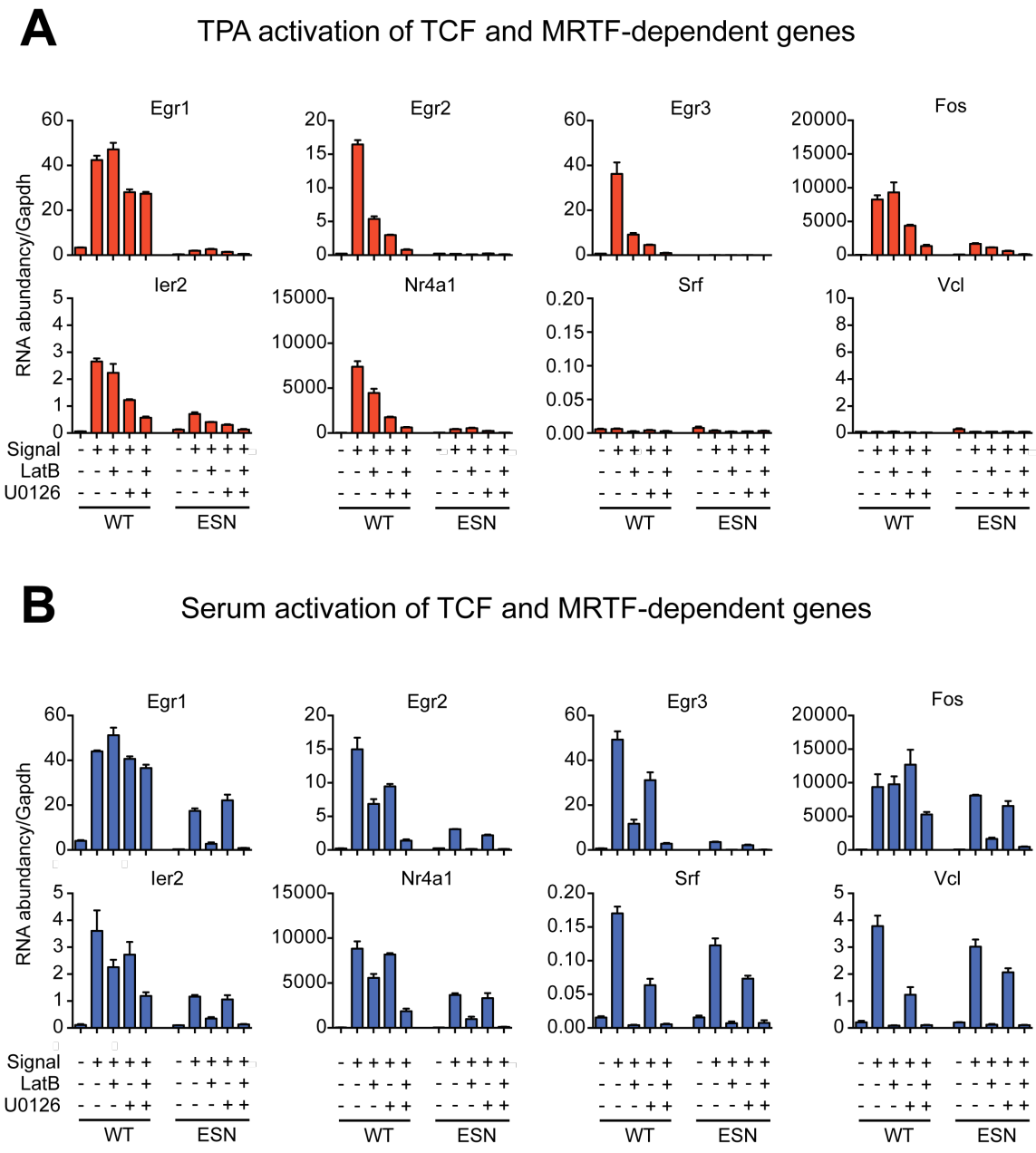


Figure 5.1 TCF-dependent gene activation in response to serum or TPA.

(A) Activation of TCF and MRTF dependent genes in response to TPA in wild type MEF and ESN. Cells were serum starved for 48h and stimulated for 30 minutes with TPA. Pre-treatment of U0126, LatB or a combination of the two was done 30 minutes before TPA stimulation**. (B)** Activation of TCF and MRTF dependent target as in panel A in response to serum stimulation. *Egr1, Egr2, Egr3, Fos, Ier2* and *Nr4a1* are TCF-dependent targets. *Srf* and *Vcl* are MRTF dependent targets. RNA was collected and quantified by RT-PCR. The values are expressed as relative to *Gadph* abundance (This data was collected by Cyril Esnault).

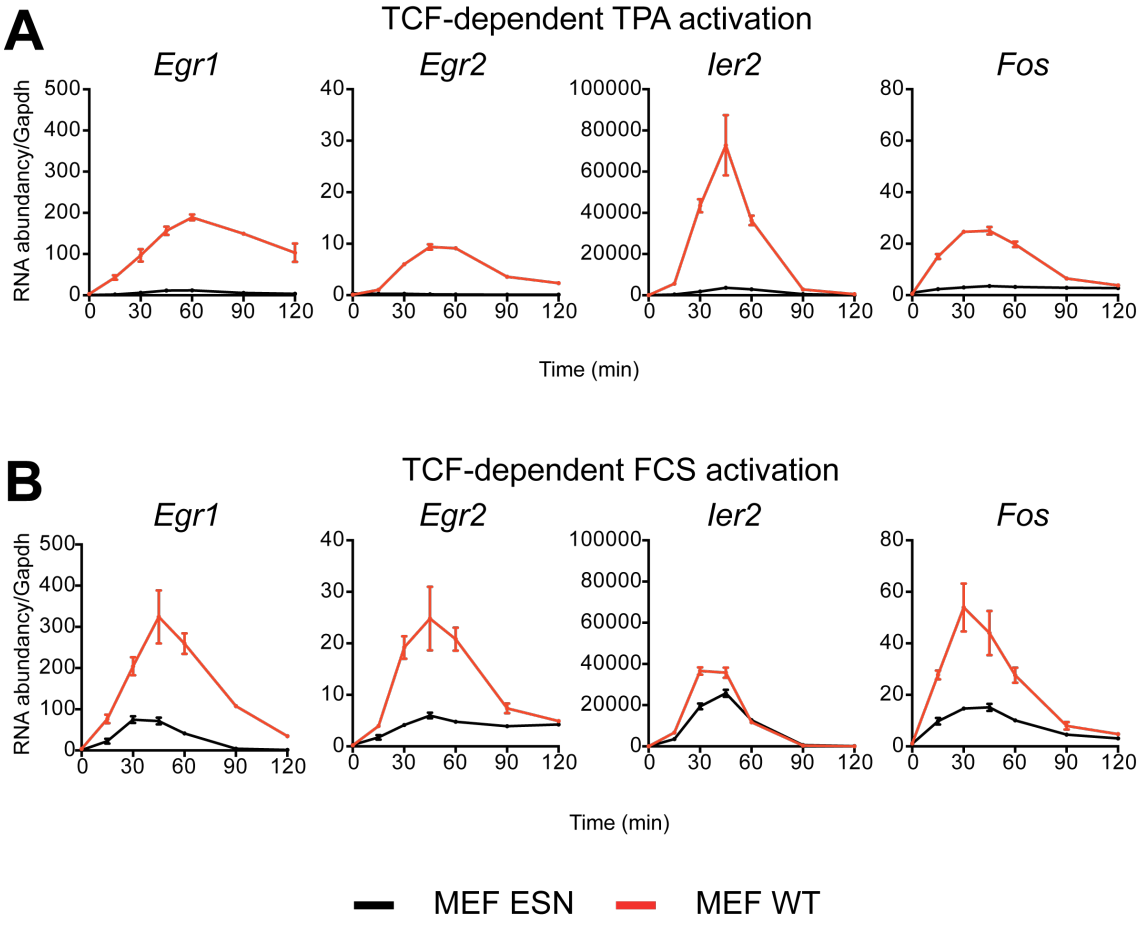


Figure 5.2 TCF-dependent IEGs activation.

(A and B) Time course stimulation of wild type or ESN MEFs with TPA (panel A) or serum (panel B) for TCF-dependent targets. Cells were serum starved for 48h and stimulated for 15, 30, 45, 60, 90 and 120 minutes. The collected RNA was quantified with RT-PCR and normalised to gapdh. Red line shows the expression in wild type MEF while the black line shows the induction in ESN MEF.

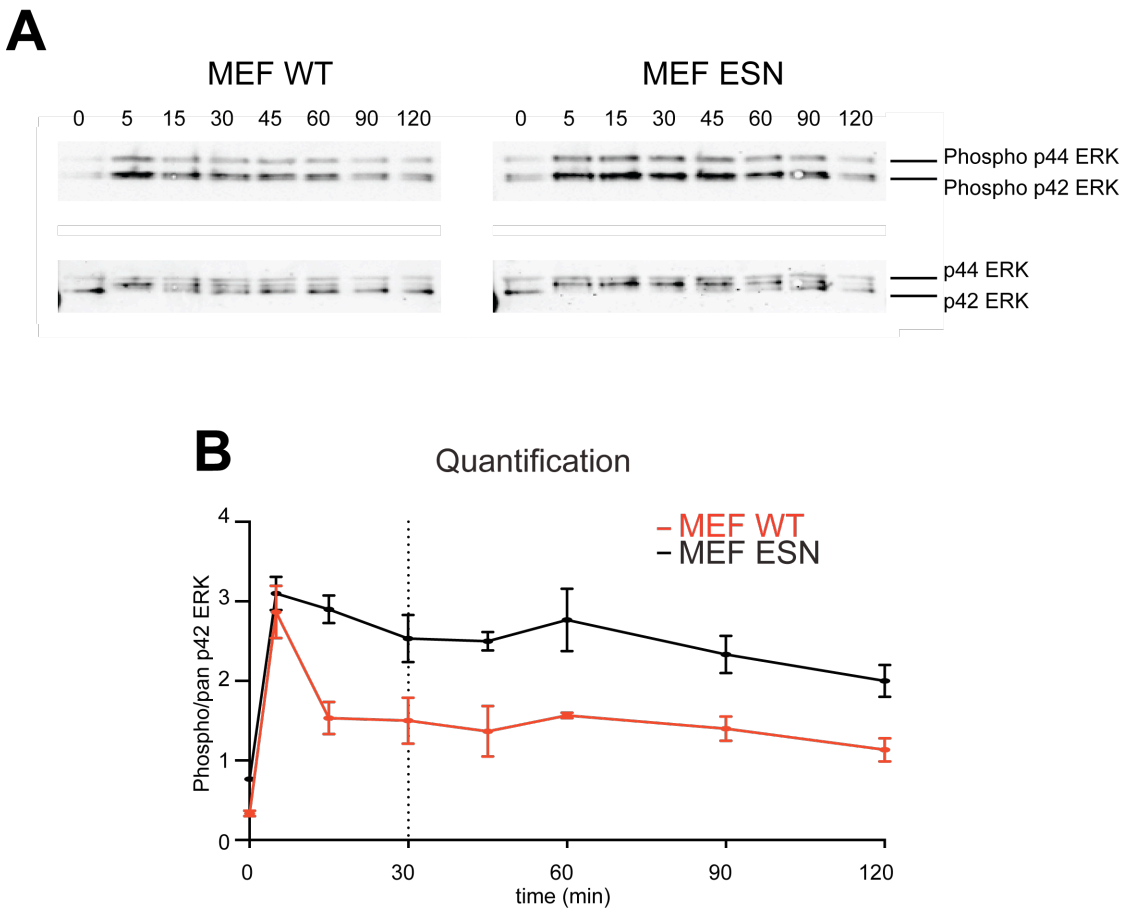


Figure 5.3 Erk1/2 activation in response to TPA in wild type and ESN MEFs.

(A) Western blot analysis of cell extract obtained from either wild type MEF (left) or ESN MEF (right) after treatment with TPA for 5, 15, 30, 45, 60, 90 and 120 minutes following 48h serum starvation. The antibodies used are phospho-p44/42 (Erk1/2) (Thr202/tyr204) (top) or control antibody p44/42 MAPK (Erk1/2) (bottom). **(B)** Quantification of the phospho-p44/42 antibody signal from a triplicate experiment as in panel A normalised to control p44/42 signal.

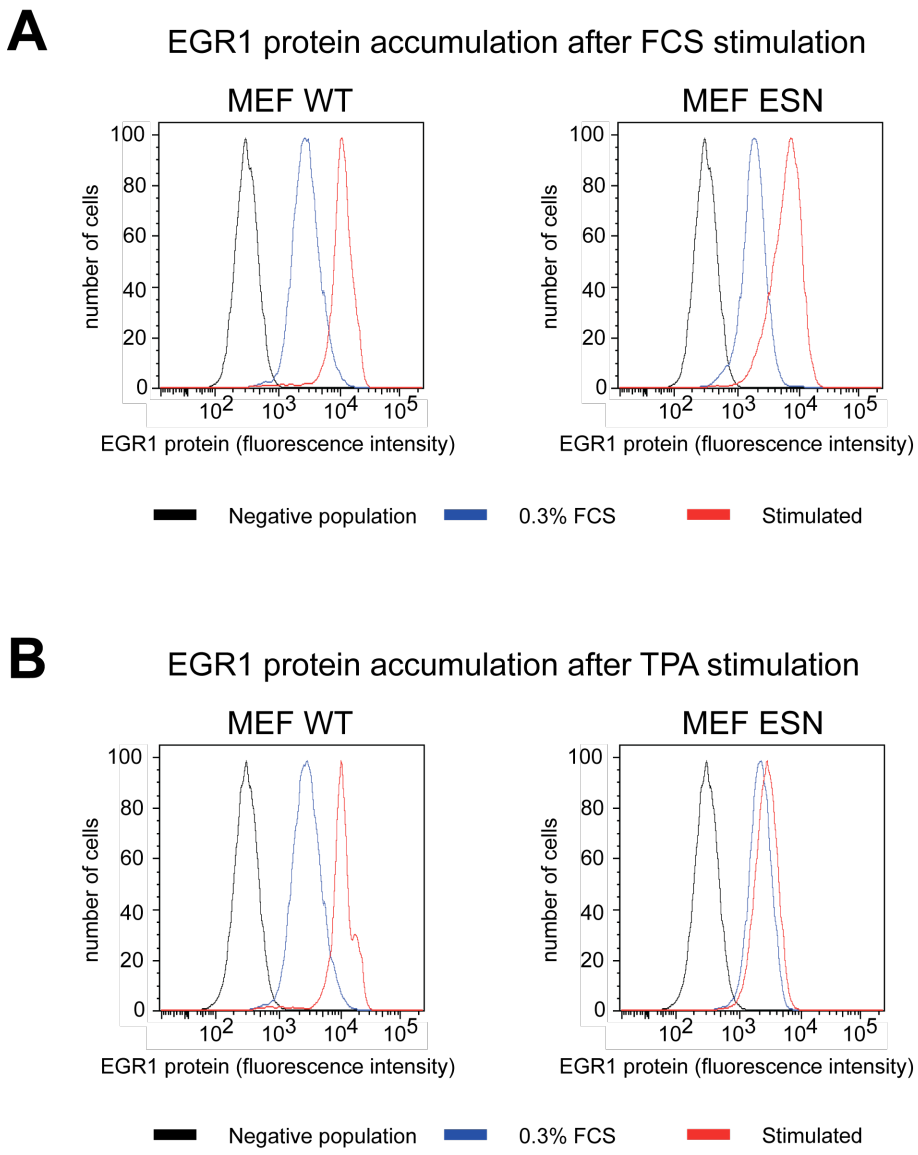


Figure 5.4 *Egr1* protein accumulation after serum or TPA in wild type or ESN MEFs.

Analysis by flow cytometry of *Egr1* protein expression after FCS (panel A) or TPA (panel B) stimulation for wild type or ESN MEFs. Cells were starved for 48h and stimulated for 1h with either FCS or TPA. Cells were ethanol fixed, permeabilised and stained with primary antibody against *Egr1* following a secondary antibody Cy3 labelled recognising the primary antibody. A negative control was included where only the secondary Cy3 labelled antibody was used (black line, 'Negative population')

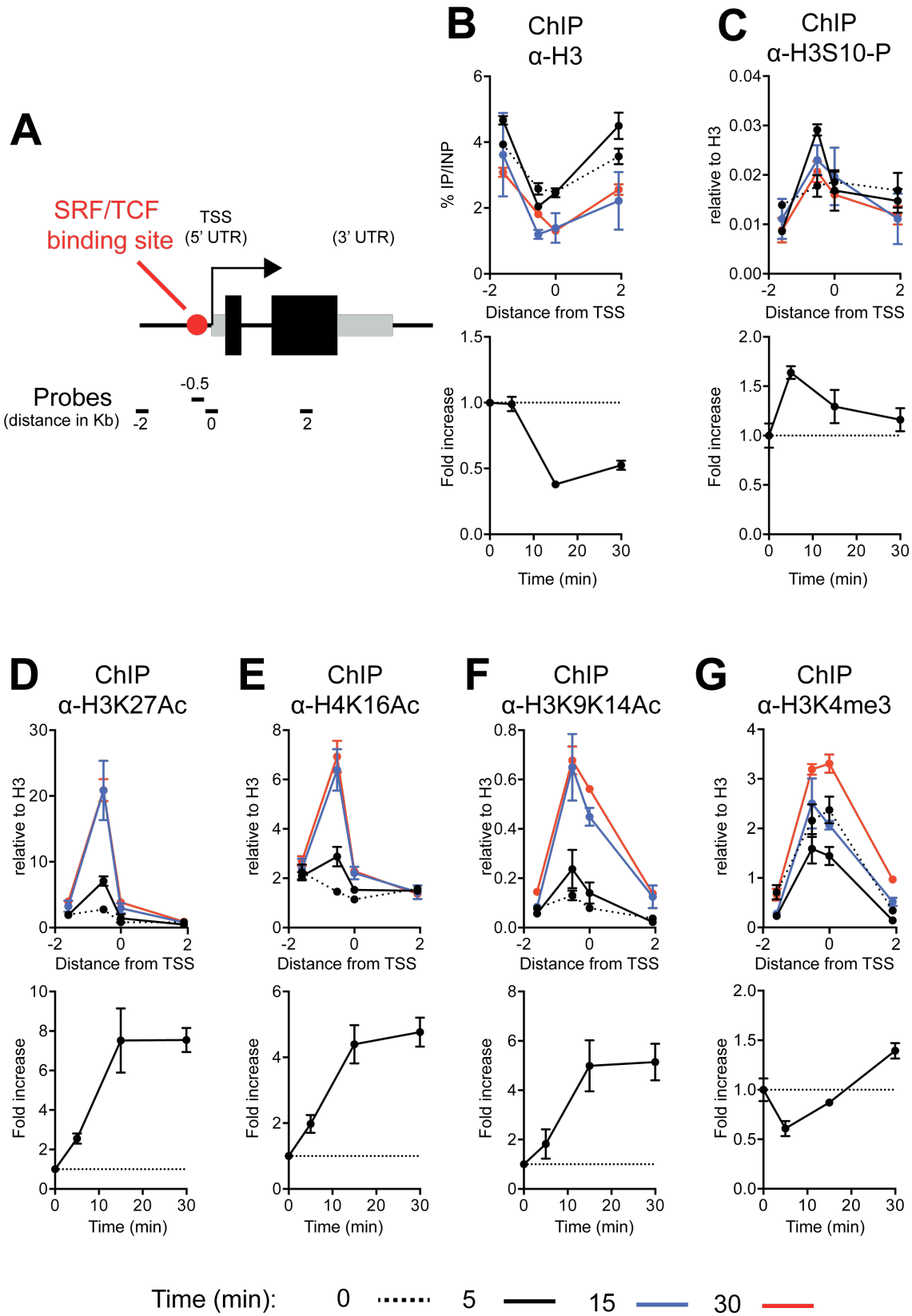

A
SRF/TCF binding site
TSS
(5' UTR)
(3' UTR)
Probes
(distance in Kb)
-2
-0.5
0
2
B
ChIP
α-H3
% IP/INP
Distance from TSS
Fold increase
Time (min)
C
ChIP
α-H3S10-P
relative to H3
Distance from TSS
Fold increase
Time (min)
D
ChIP
α-H3K27Ac
E
ChIP
α-H4K16Ac
F
ChIP
α-H3K9K14Ac
G
ChIP
α-H3K4me3
relative to H3
Distance from TSS
Fold increase
Time (min)
Time (min): 0 5 15 30

Figure 5.5 Establishment of a chromatin signature at *Egr1* promoter in response to TPA.

(A) Representation of the *Egr1* gene illustrating the positions of different probes along the gene. Exons are indicated with 5'UTR and 3'UTR in grey and coding sequences in black; SRF binding site is indicated with a red circle. **(B to G)** ChIP experiments using the indicated antibody. At the top of each panel is the profile at *Egr1* gene while at the bottom is the kinetic of the signal at the coordinate showing the best signal (for Total H3 at the TSS, for H3S10-P, H3K27Ac, H4K16ac and H3K9K14ac at the SRF binding site while for H3K4me3 at the *Egr1* TSS).

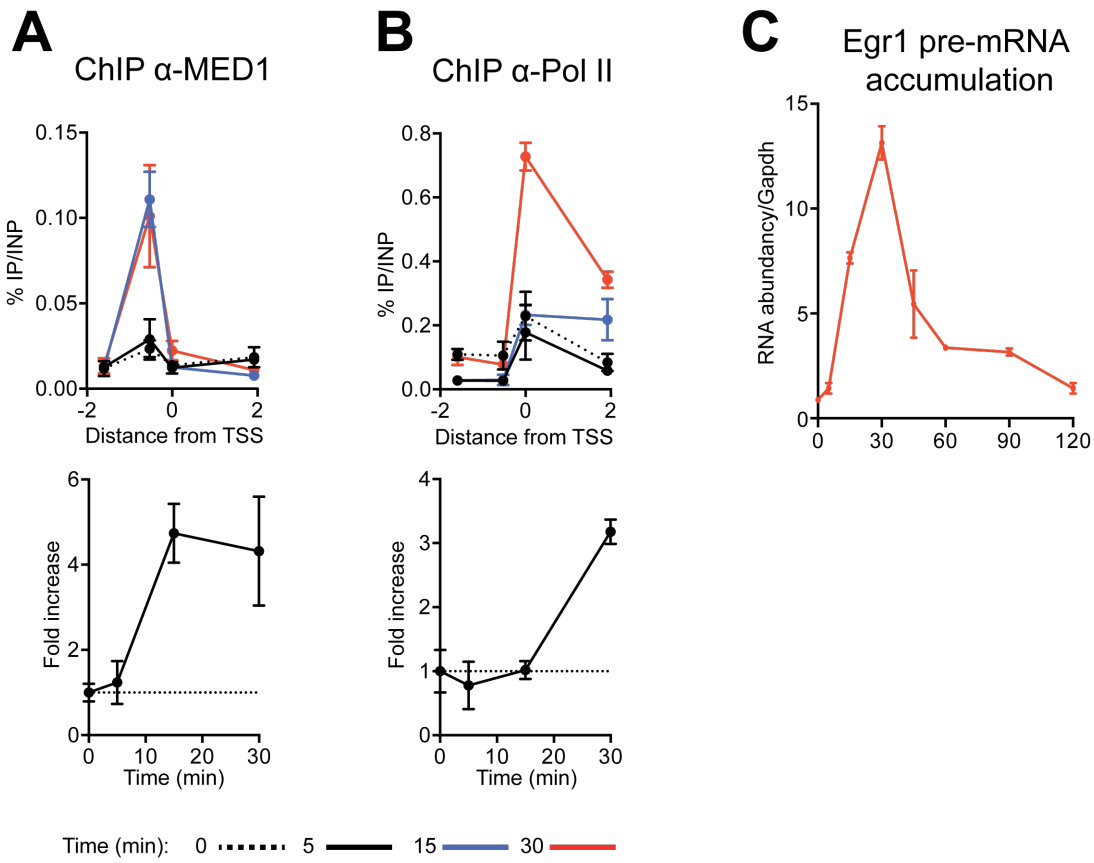


Figure 5.6 Med1 and Pol II recruitment following TPA stimulation at *Egr1* promoter correlates with pre-mRNA accumulation.

(A and B) ChIP time course of Med1 (panel A) and Pol II (panel B). At the top of each panel is the profile at *Egr1* gene while at the bottom is the kinetic across the time course of the signal at the coordinate showing the best signal (for Med1 at the SRF binding site while for Pol II at *Egr1* TSS). The coordinates on the x-axes refer to Figure 4.5A. **(C)** *Egr1* pre-RNA accumulation over the same time course and including 45, 60, 90 and 120 minutes following TPA stimulation.

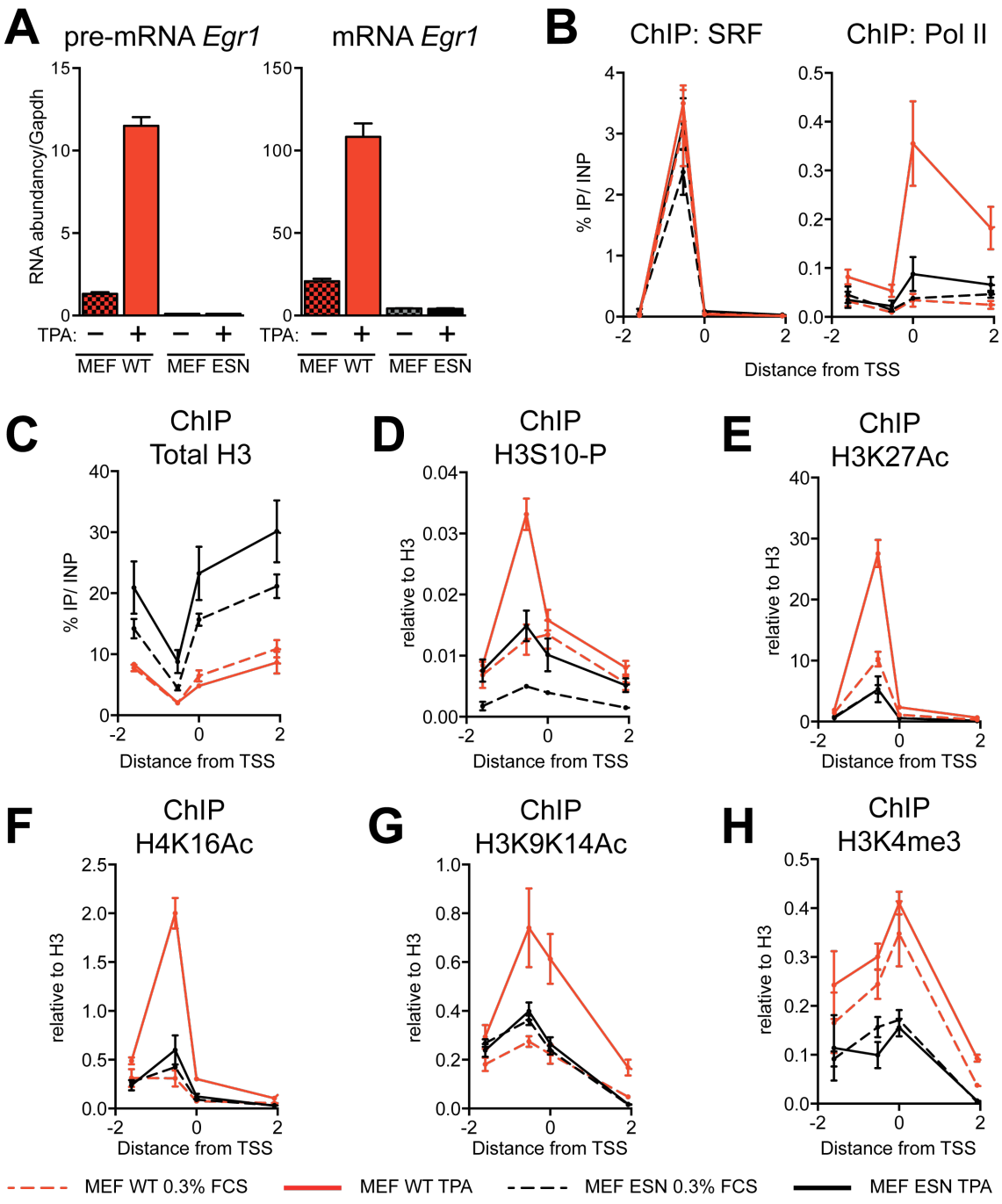
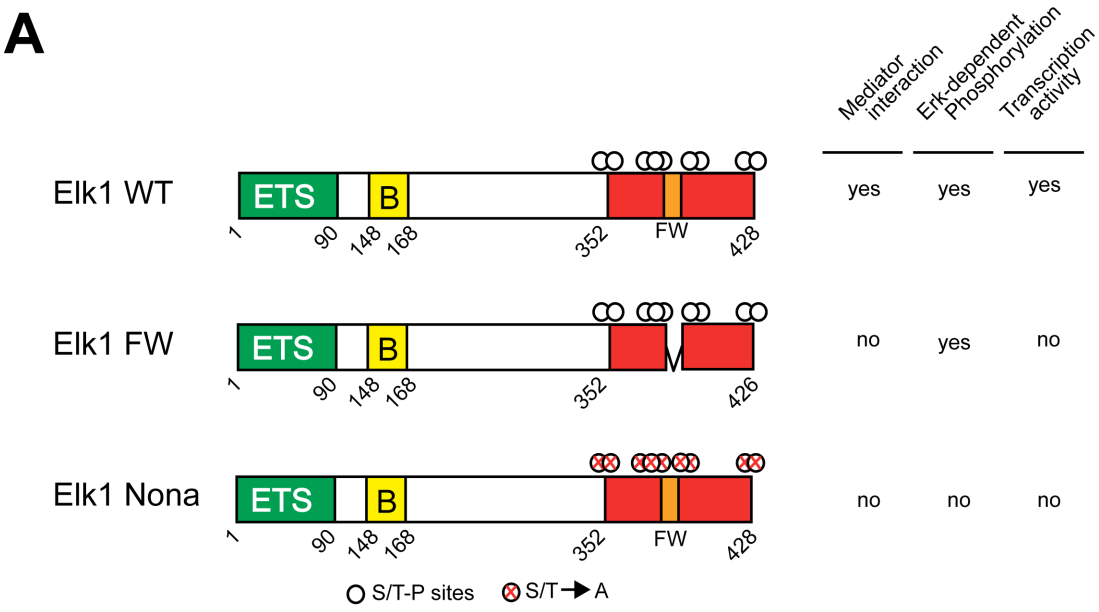


Figure 5.7 *Egr1* chromatin signature requires the TCFs.

(A) Expression of pre-RNA (left) and mRNA (right) of *Egr1* after 30 minutes TPA stimulation in wild type MEF (red) and ESN MEF (black). **(B to H)** ChIP using the indicated antibodies at *Egr1* gene. Dashed lines indicate resting condition while the continuous line shows 30 minutes TPA stimulation. In red is the wild type MEF profile while in black are the ESN MEF profiles. The coordinates on the x-axes refer to Figure 4.5 A.


A
Mediator interaction
Erk-dependent Phosphorylation
Transcription activity
Elk1 WT
ETS
B
1
90
148
168
352
FW
428
yes
yes
yes
Elk1 FW
426
no
yes
no
Elk1 Nona
no
no
no
S/T-P sites
S/T→A

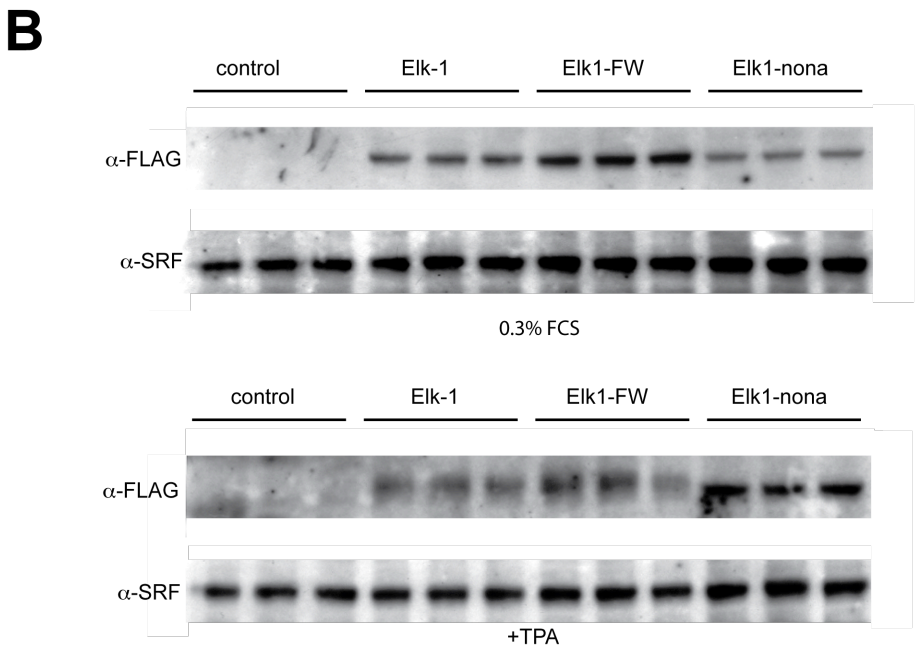

B
control
Elk-1
Elk1-FW
Elk1-nona
α-FLAG
α-SRF
0.3% FCS
+TPA

Figure 5.8 Establishment of reconstituted ESN MEF cell lines.

(A) Schematic representation of the mutants used in the reconstitution. Human Elk-1 is a 428 amino acid polypeptide consisting of three domains: ETS DNA-binding domain (green) at the N-terminal part; a B-box involved in SRF interaction (yellow); activation domain (red) harbouring S/T-P sites and two hydrophobic residues FW. The ability of each mutant to interact with Mediator, being phosphorylated by Erk and to activate transcription is reported on the right hand side of the panel. **(B)** Western blot of protein extract obtained for each reconstituted cell line in serum starved (top) and TPA stimulated conditions (bottom). SRF was used as a loading control while an anti-FLAG antibody was used to show the exogenous proteins. The samples are obtained from Chromatin preparations later used for ChIP experiments. Each replicate correspond to the one used in ChIP experiments.

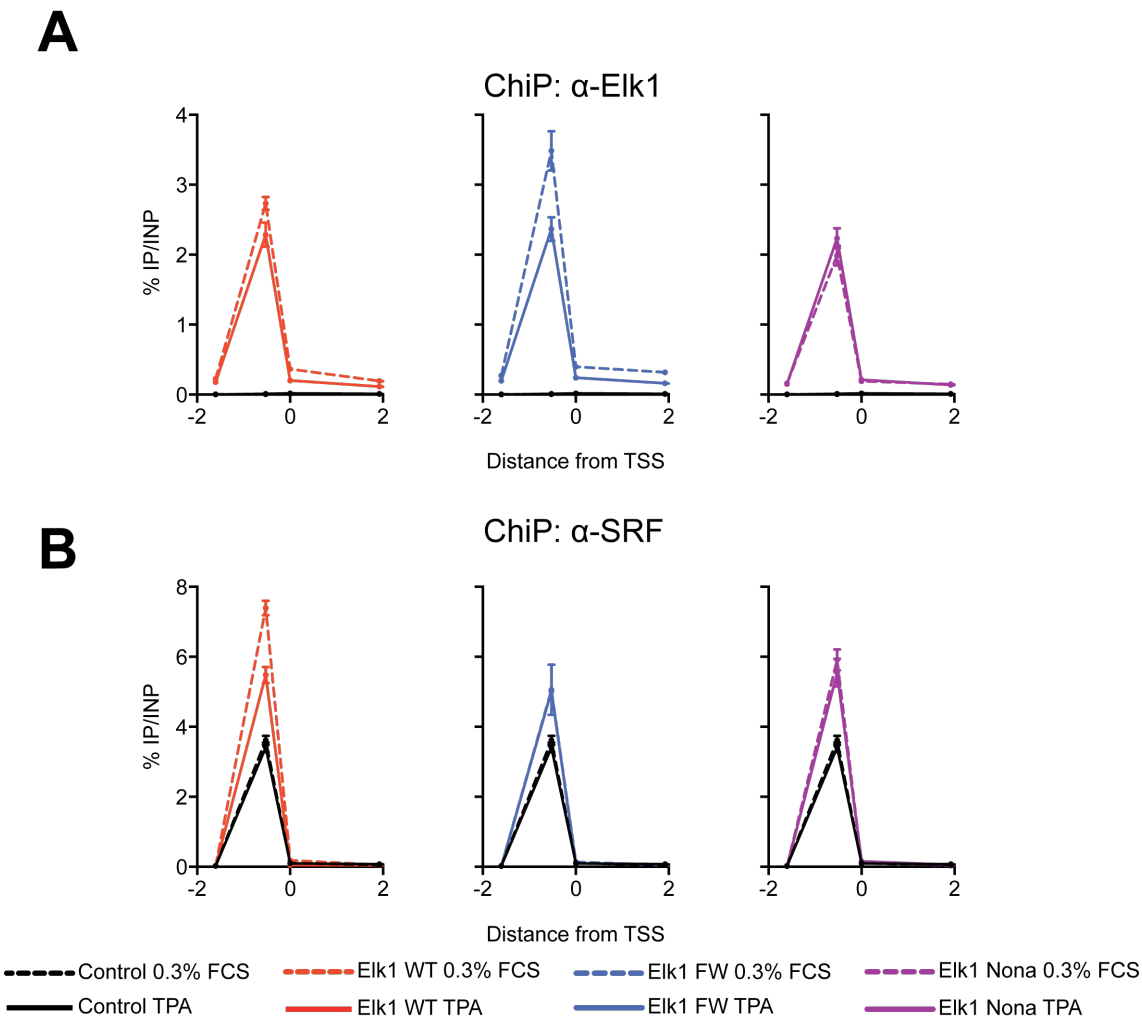


Figure 5.9 Elk-1 and SRF binding at *Egr1* promoter in reconstituted cell lines.

ChIP using an anti-Elk-1 antibody **(A)** and an anti-SRF antibody **(B)** across cells expressing different Elk-1 derivatives and vehicles (black line). Dashed lines represent serum-starved cells while continuous lines represent TPA stimulated conditions. In red are ESN MEF infected with retrovirus expressing Elk-1 wild type, in blue with Elk-1 FW and in purple with Elk-1 Nona. The coordinates on the x-axes refer to Figure 4.5 A.

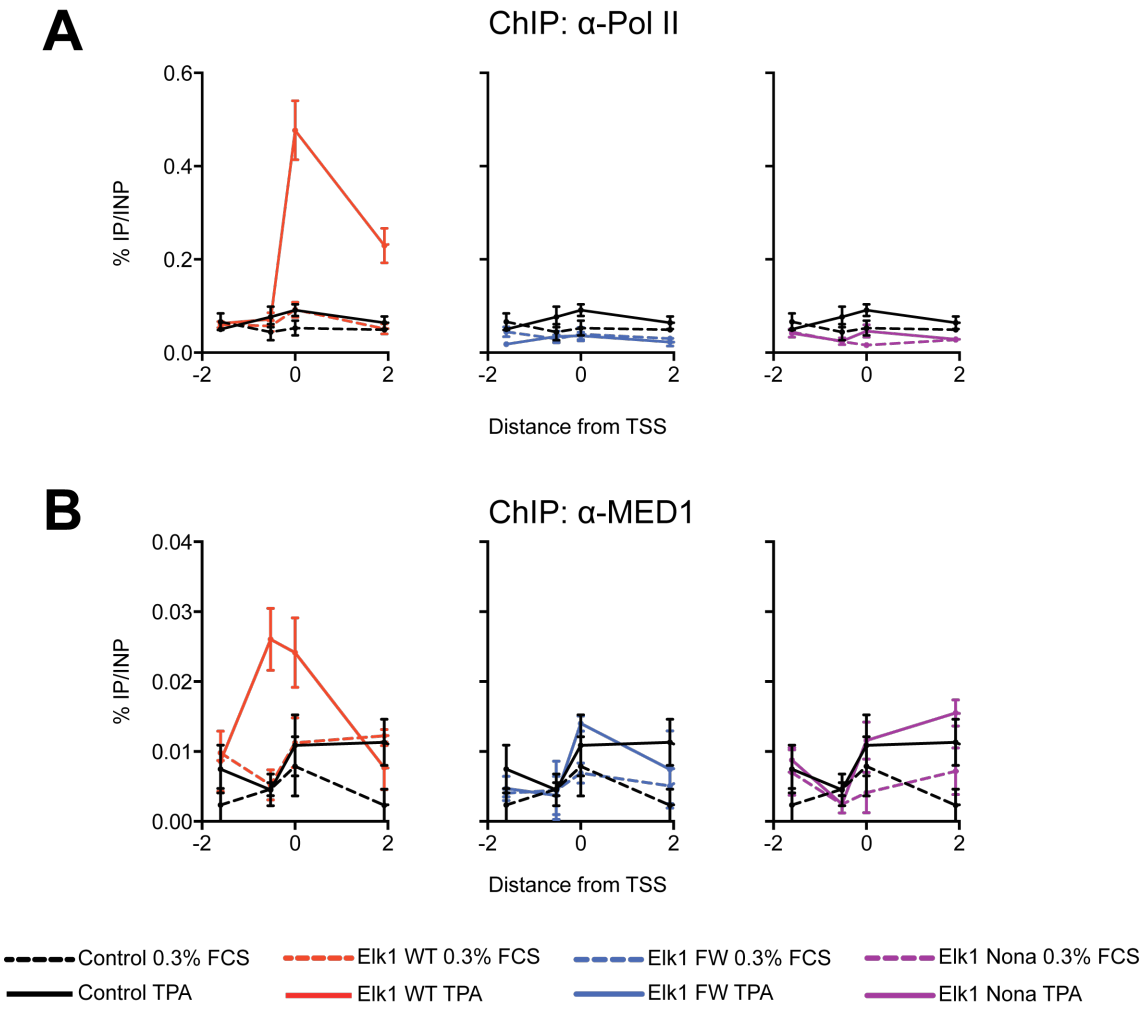


Figure 5.10 Elk-1 AD coordinates Mediator and Pol II recruitment at *Egr1* promoter.

ChIP experiments using an anti-Pol II (8WG16) **(A)** and an anti-Med1 antibody **(B)** across the reconstituted cell lines. The coordinates on the x-axes refer to Figure 4.5 A.

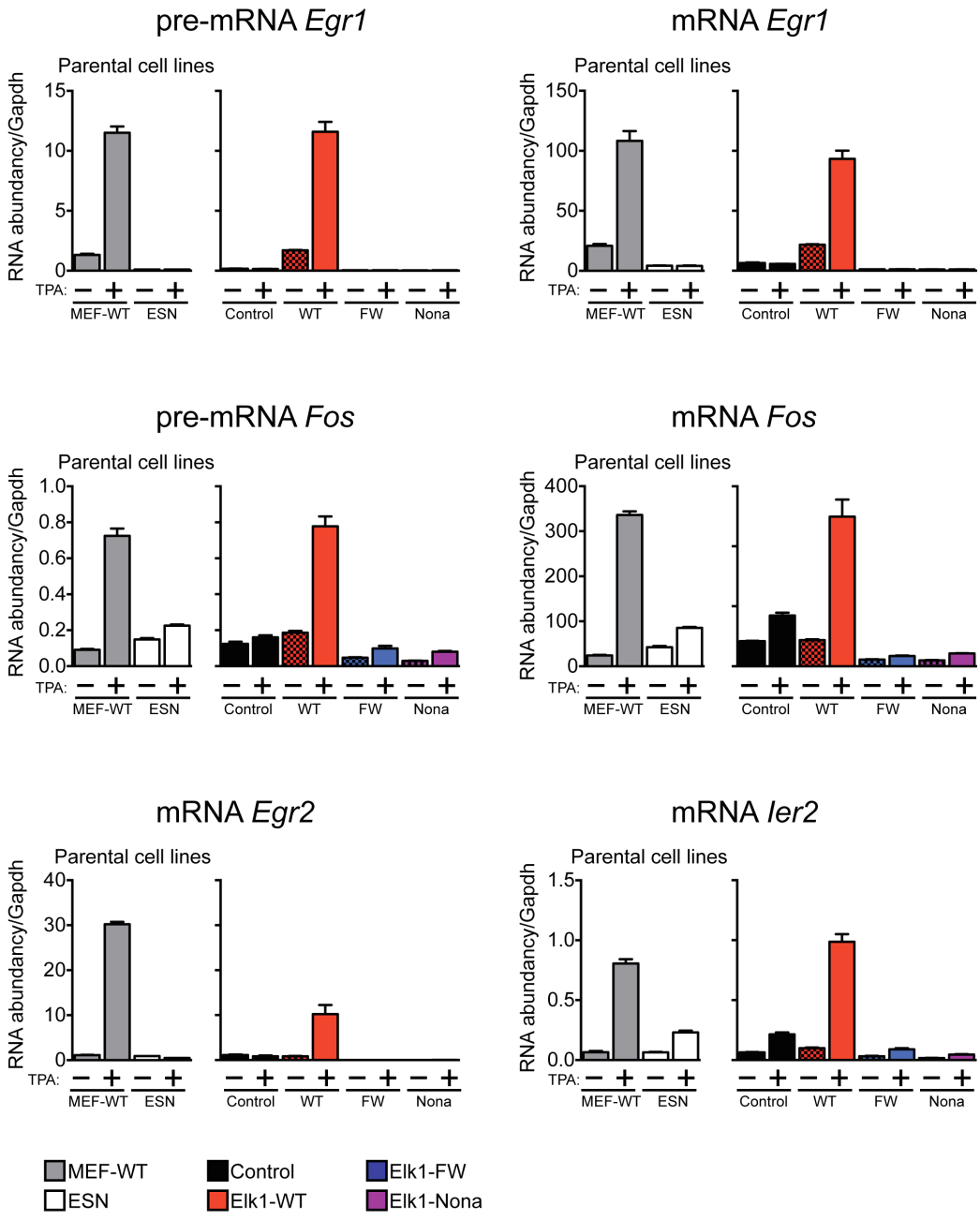


Figure 5.11 SRF/TCF-target genes expression in reconstituted cell lines.

Expression analysis of reconstituted cell lines after TPA stimulation. At the left hand side of each panel is the expression level in parental cell lines: Wild type MEF (grey) and ESN MEF (white). On the right hand side of each panel the expression of target genes is shown in reconstituted ESN MEF: control infection (black), Elk-1 wild type (red), Elk-1 FW (blue), Elk-1 Nona (purple). The targets analysed are *Egr1* at pre-RNA and mRNA, *Fos* at pre-RNA and mRNA, *Egr2* and *Ier2* at the mRNA level.

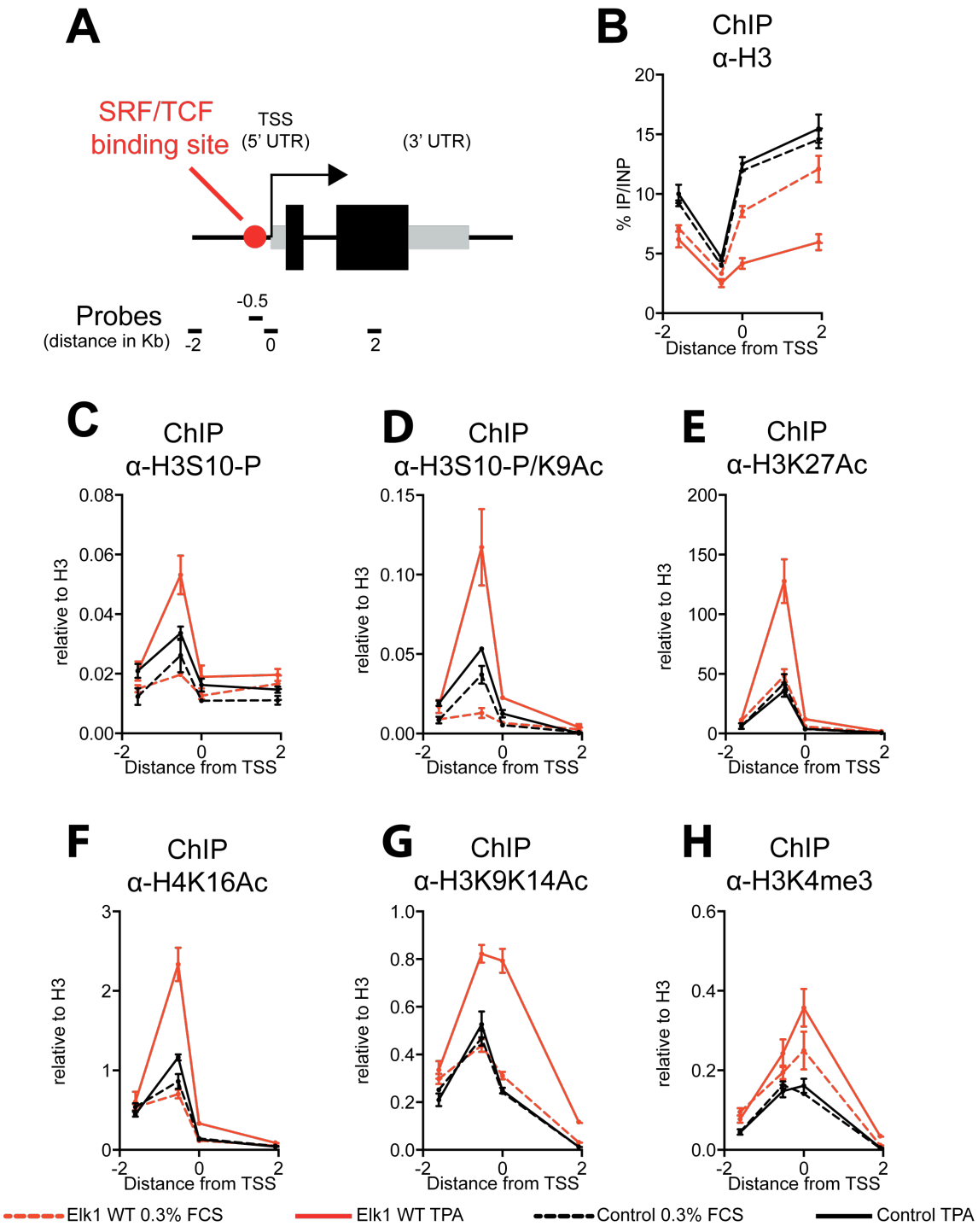


Figure 5.12 Elk-1 wild type is sufficient to re-establish chromatin changes at *Egr1* promoter.

(A) Representation of the *Egr1* gene illustrating the positions of different probes along the gene. Exons are indicated with 5'UTR and 3'UTR in grey and coding sequences in black; SRF binding site is indicated with a red circle. **(B to H)** ChIP experiments using the indicated antibody.

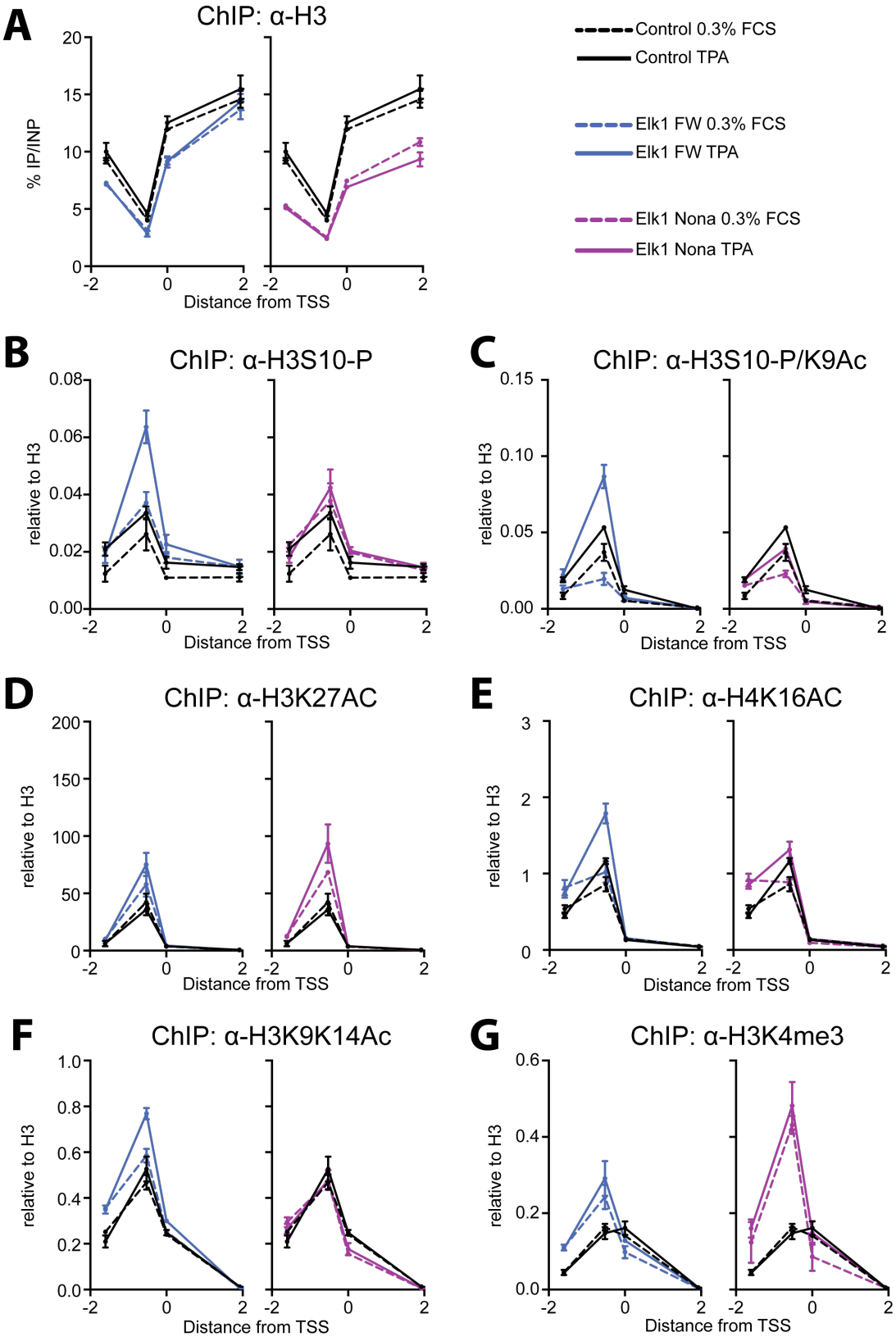
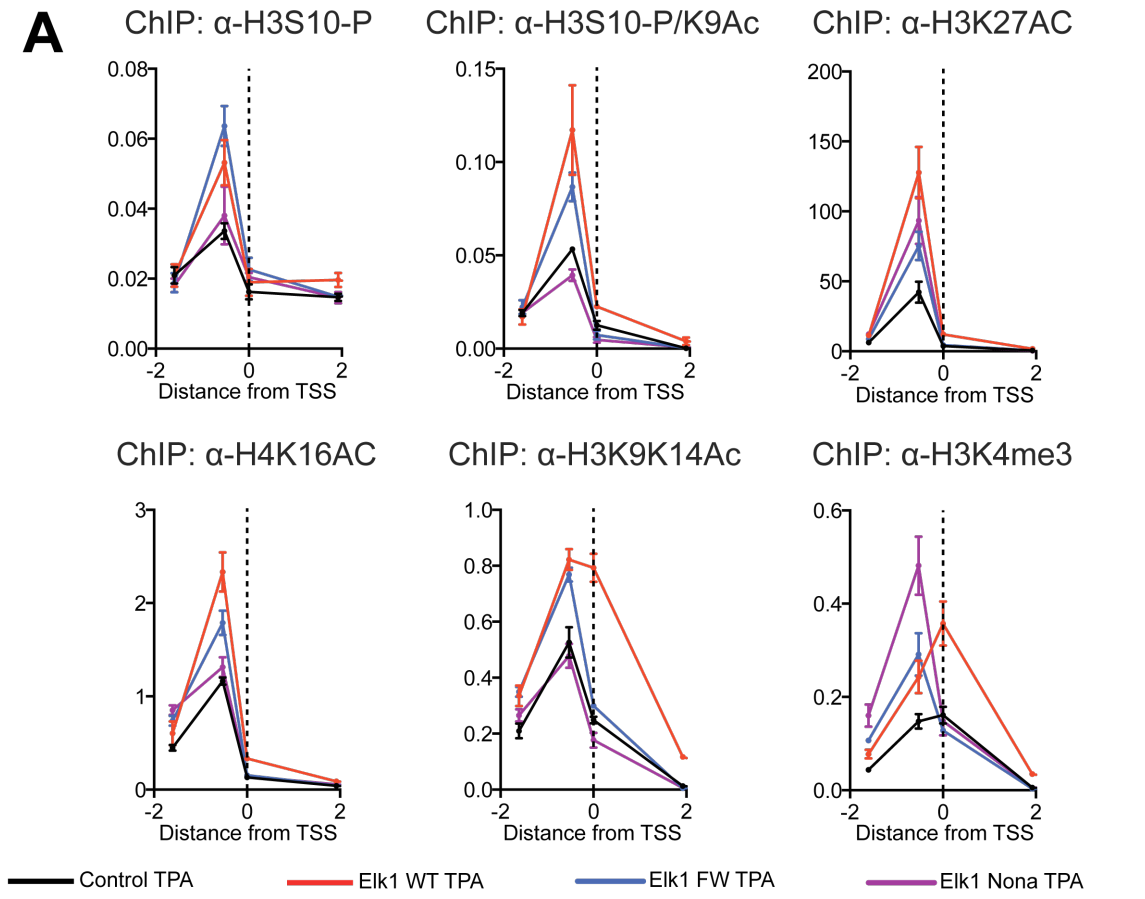

A
ChIP: α-H3
% IP/INP
Distance from TSS
Control 0.3% FCS
Control TPA
Elk1 FW 0.3% FCS
Elk1 FW TPA
Elk1 Nona 0.3% FCS
Elk1 Nona TPA
B
ChIP: α-H3S10-P
relative to H3
C
ChIP: α-H3S10-P/K9Ac
D
ChIP: α-H3K27AC
E
ChIP: α-H4K16AC
F
ChIP: α-H3K9K14Ac
G
ChIP: α-H3K4me3

Figure 5.13 Elk-1 DNA binding and activation domain activities contribute to defined chromatin changes in absence of transcription.

(A to G) ChIP experiments using the indicated antibody in ESN MEF reconstitute with either Elk-1 FW (blue) or Elk-1 Nona (purple) across the *Egr1* gene. The coordinates on the x-axes refer to Figure 4.12 A.


A
ChIP: α-H3S10-P
ChIP: α-H3S10-P/K9Ac
ChIP: α-H3K27AC
ChIP: α-H4K16AC
ChIP: α-H3K9K14Ac
ChIP: α-H3K4me3
Distance from TSS
Control TPA
Elk1 WT TPA
Elk1 FW TPA
Elk1 Nona TPA

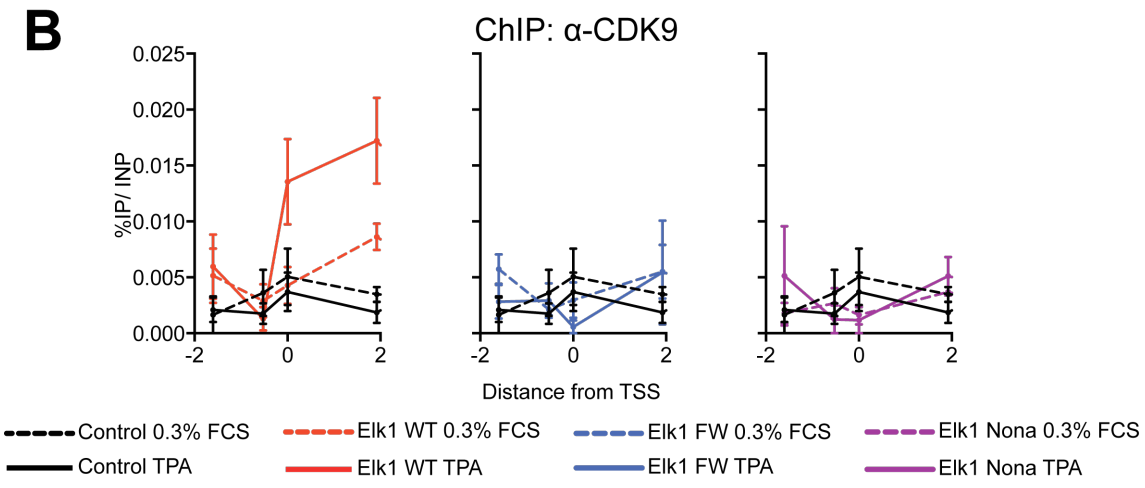

B
ChIP: α-CDK9
%IP/ INP
Distance from TSS
Control 0.3% FCS
Elk1 WT 0.3% FCS
Elk1 FW 0.3% FCS
Elk1 Nona 0.3% FCS
Control TPA
Elk1 WT TPA
Elk1 FW TPA
Elk1 Nona TPA

Figure 5.14 Chromatin changes are not sufficient for p-TEFb recruitment.

(A) Comparison of the ChIP signal across *Egr1* promoter of reconstituted ESN MEFs with control (black), Elk-1 wild type (red), Elk-1 FW (blue) and Elk-1 Nona (purple) for the TPA stimulated condition. The data presented is as in Figure 5.13. The antibodies used are shown at the top of each graph. Dotted line shows the TSS position of the *Egr1* gene. **(B)** ChIP of CDK9 at *Egr1* gene in ESN MEF reconstituted with Elk-1 derivative using the same colour code as in panel A. The dashed line shows 0.3% FCS condition while the continuous line shows the TPA stimulated conditions

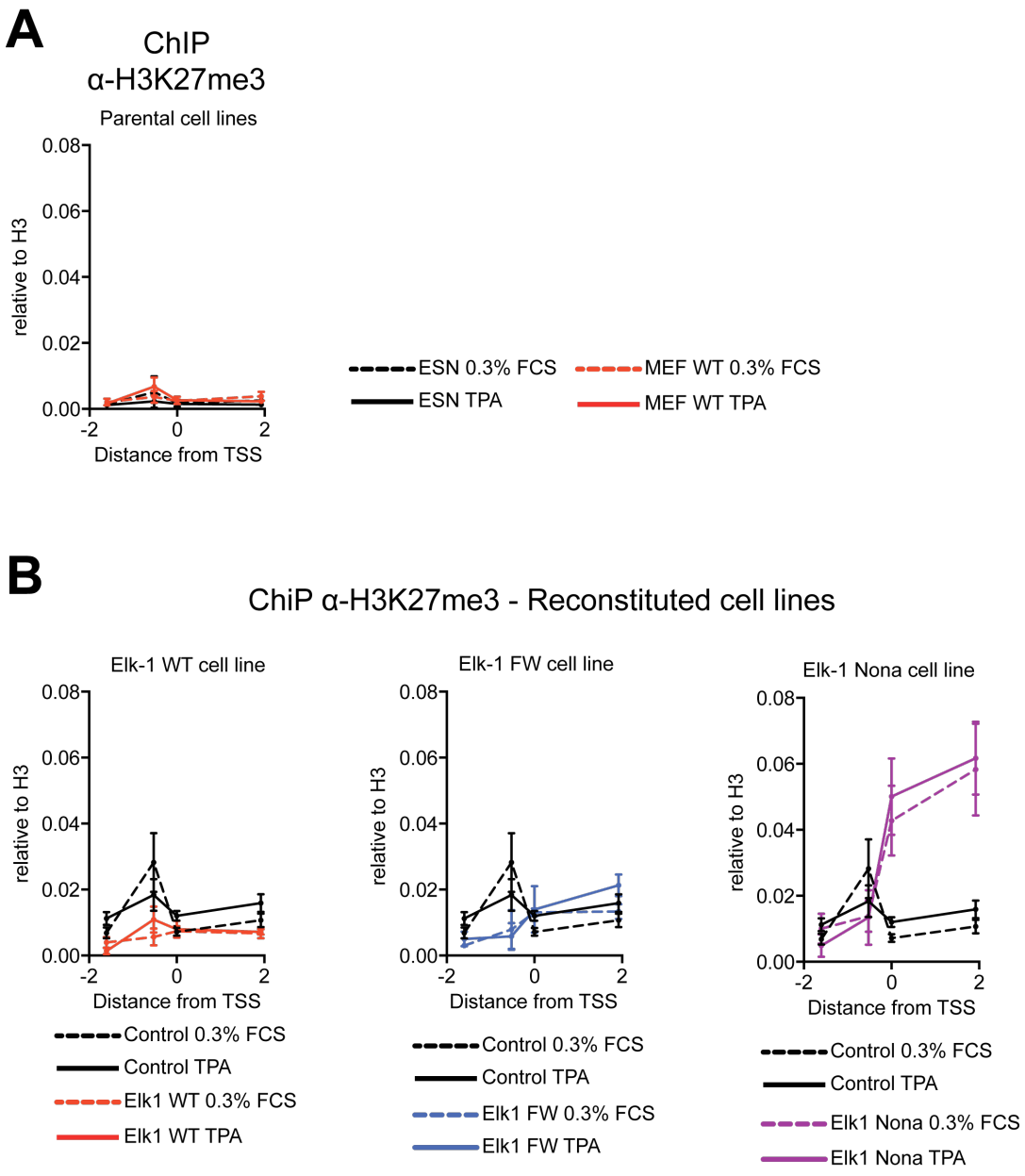


Figure 5.15 Elk-1 activation domain is required to maintain a permissive chromatin at *Egr1* ORF.

(A) ChIP using an H3K27me3 antibody in Wild type MEF (red) or ESN MEF (black). The signal is shown relative to the signal obtained with anti-H3 (see Figure 5.7, 5.12 and 5.13). **(B)** ChIP using an H3K27me3 antibody in ESN MEF reconstituted with Elk-1 wild type (red), Elk-1 FW (blue) and Elk-1 Nona (purple). The signal is shown relative to the signal obtained with anti-H3 (see Figure 5.12 and 5.13).

Chapter 6. TCF-dependent activation requires a defined set of chromatin remodellers and modifiers

6.1 Aim

TCF-dependent activation, as described in Chapter 5, involves a defined set of chromatin changes that rely on the TCFs. Chromatin modifications per-se are not sufficient to induce transcription while a correct interplay between signal-induced chromatin changes and the transcriptional machinery is mediated by the TCF AD. In order to characterise the role of the observed chromatin modifications we aimed to identify the modifiers and remodellers involved in TCF-dependent gene activation. The data presented in this chapter is preliminary as we are currently collecting further data in order to validate these observations. I am going to present what has so far been analysed.

6.2 Medium-throughput siRNA screen and validation

To identify chromatin modifiers involved in immediate-early gene expression we used a medium-throughput siRNA approach. To reduce complexity, we focused on catalytically active proteins amongst the ~1000 chromatin-modifiers and remodellers (Fazzio et al. 2008). We selected 50 genes including 13 methyltransferases, 14 acetyltransfearses, 7 kinases, 5 phosphatases and 9 remodellers (Figure 6.1A). The siRNA screen also included two control oligos, with no target in mouse cells, and two positive controls such as Srf and Med23, known to be essential for TCF-dependent gene activation. The siRNA screen was performed in wild type MEF harbouring all three TCFs.

The initial phase of the screen was performed using the Dharmacon ON-target plus siRNA pools for each of the 50 targets. Each pool comprises four independent siRNA duplexes against the target genes. Four readouts, consisting of pre-RNA and mature RNA of both *Egr1* and *Fos* gene, were monitored by quantitative PCR. The knockdown of Srf was used as parameter for siRNA efficacy and as a threshold for significant reduction in gene expression after TPA stimulation.

An On-target pool was considered a hit if it was impairing at least 3 out of 4 readouts, with a reduction of at least 20% (minimum reduction observed for Fos mRNA after Srf knockdown) (Figure 6.2). The expression of *Egr1* and *Fos* showed a good correlation at both pre-mRNA and mature RNA (Figure 6.1B). A higher fluctuation was observed at the pre-RNA level reflecting its reduced stability if compared to mRNA of both targets. This initial screen allowed us to select 13 potential candidates (Figure 6.2).

To validate the hits selected from the initial phase, a deconvolution secondary screen was performed in which the four duplexes, from each of the 13 hits were individually screened at a final concentration of 20nM (Figure 6.3). We considered hits those that showed at least two out of four oligos impairing at least 3 out of 4 readout (using the same threshold as in phase 1) (Figure 6.3). We selected a final list of 7 targets shown to impair the TPA response of both *Egr1* and *Fos*. This refined group included the methyltransferase MLL3, known to methylate lysine 4 of the histone H3 (H3K4) (Y.-W. Cho et al. 2007); the chromatin remodeller CHD2, known to interact with H3K4me3 and deposit histone H3.3 at the TSS of active genes (J. A. Hall & Georgel 2007); the methyltransferase Set7 (KMT7), known to mono-methylate H3K4 and associate with Tat proteins to stimulate Pol II elongation (Rice 2010); the SET domain-containing protein SMYD2 (KMT3C), reported to methylate H3K4 and demethylate H3K36me2 (Brown et al. 2006); the helicase Ruvbl2 (Tip49b), part of the Nua4 complex involved in histone acetylation and chromatin remodelling in conjunction with the SWR1-like complex (Kanemaki et al. 1999); the acetyltransferase Tip60 (KAT5), shown to specifically acetylate Lysine 16 of the histone H4 (K. K. Lee & Workman 2007); and the kinase Aurora-B (AURKB) that phosphorylates histone protein during cell division and recently was shown to associate to defined promoters in quiescent lymphocytes (Frangini et al. 2013).

6.3 Aurora-B is specifically required for TCF-dependent gene activation

The siRNA screen we performed provided us with a limited number of potential modifiers and remodellers involved in TCF-dependent gene activation. We initially focused on the validation of Aurora-B as for us it was the most unexpected

hit in the screen. Aurora proteins are kinases essential for cell proliferation, controlling chromatid segregation by phosphorylating H3S10 residues leading to HP1 disassociation from heterochromatin during mitosis (Fischle et al. 2005). Aurora-A is also involved in cell proliferation and cell division, therefore in order to discount negative effects in transcriptional activity we compared the knockdown of both Aurora A and B and their effects in *Egr1* and *Fos* activity. Only Aurora-B knockdown was specifically affecting *Egr1* and *Fos* expression after TPA stimulation (Figure 6.4 A). Strikingly this impairment was observed only at TCF-dependent targets, as down-regulation of Aurora-B did not affect *Acta2* expression (an SRF-MRTF-specific target) after CD or serum stimulation (Figure 6.4 B).

In order to exclude long-term effects caused by impairing Aurora-B expression we selected two compounds affecting Aurora-A and Aurora-B activities (Figure 6.4 C). Cells were treated for 5 minutes with each compound before treating for 30 minutes with TPA. Most of these compounds have a wide spectrum of action, as they are designed to compete for ATP in the active site of the target kinase. Hesperidin, an Aurora-B specific drug with IC50 of 250nM and no effect on MAPKs (Hauf et al. 2003), reduced of 50% the induction in both *Egr1* and *Fos* expression following TPA (Figure 6.4C). This observation corroborates a direct involvement of Aurora-B in TCF-dependent gene activation. The compound known to affect only Aurora-A did not show significant effect.

6.4 Preliminary dissection of signal-induced chromatin modification at *Egr1* promoter

In order to investigate the role of the selected modifiers and remodellers in TCF-dependent gene activation we performed ChIP experiments following down-regulation of each separate modifier and remodeller selected from the siRNA screen.

This assay was optimised with SRF and MED23 siRNAs against a non-targeting siRNA as a control. It was possible to observe a down-regulation of *Egr1* expression following SRF and MED23 knockdown (Figure 6.5A). Reduction of *Egr1* expression correlated with a reduction of SRF and Pol II ChIP signal for the SRF knockdown (Figure 6.5C and D). On the other hand MED23 knockdown did not affect SRF ChIP signal while it reduced Pol II signal at the *Egr1* promoter (Figure

6.5C and D). Most chromatin changes were left unaffected after knocking down SRF (Figure 6.5E to H). SRF knockdown seemed detrimental for the TPA-dependent H3K4me3 induction (Figure 6.5H). This observation suggests that the contributions of the TCFs and SRF for *Egr1* activation are separable but both required for active transcription. TCFs are capable of interacting with genomic loci autonomously. The ETS binding motif, CAGGAT, initially found at the Fos promoter is degenerated at the *Egr1* promoter. In particular the ETS motif CCGGAA at the *Egr1* promoter could allow autonomous TCF DNA binding in the absence of SRF (Treisman 1994). Assessment of Elk-1, Sap-1 or Net binding in absence or reduced SRF levels could provide further insight. Is still important to consider the loss of H3K4me3 observed following SRF knockdown. This effect, together with the distribution of H3K4me3 mostly at the TSS, suggests that in a wild type context the transcriptional machinery might primarily control the distribution of this mark.

Having established the assay we proceeded with impairing the remodellers and modifiers selected with the siRNA screen. Only four out of the seven hits showed a clear effect on the chromatin signature. In particular Aurora-B knockdown had a clear effect on H3S10PK9ac (Figure 6.6 C). Despite a clear impairment of *Egr1* pre-RNA accumulation, it was possible to observe only a 20% reduction in Pol II recruitment (Figure 6.6A and B). H3K27 and H3K9K14ac enhancement following TPA were also partially affected (Figure 6.6D and E). In addition it was possible to observe a clear effect on H3K4me3 mostly at the *Egr1* TSS (Figure 6.6F)

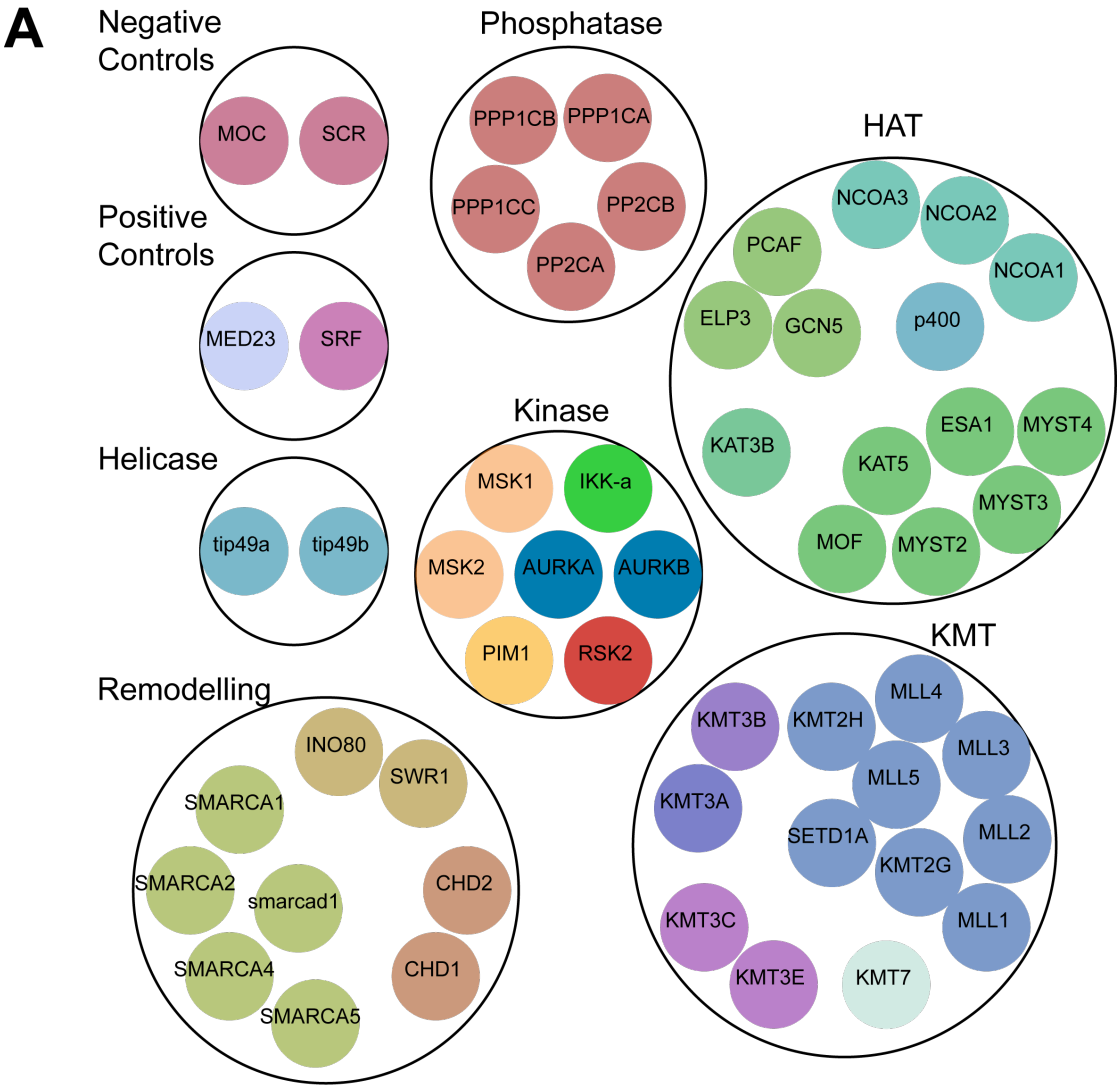
MLL3, CHD2 and SMYD2 knockdown were shown to impair few marks at the *Egr1* promoter (Figure 6.7). KAT5, Tip49b and SET7 did not show significant changes. MLL3 and CHD2 only mildly impaired Pol II recruitment (10-20% reduction at the *Egr1* TSS) (Figure 6.7B). A marked reduction in H3K4me3 when either MLL3 or CHD2 were knocked down was observed, consistent with their presumed activity (Figure 6.7F). Furthermore MLL3 and CHD2 knockdown were also shown to impair H3S10PK9ac and H3K27ac and only partially H3K9K14ac (Figure 6.7C, D and E). SMYD2 knock down did not affect most of the chromatin modifications analysed (Figure 6.8). H3K4me3 was the only modification that was impaired by SMYD2 knock down following TPA induction at *Egr1* TSS (Figure 6.8F).

Although only preliminary, these data suggest that Aurora-B, MLL3, CHD2 and SMYD2 are playing a role in establishing chromatin modifications at the *Egr1* promoter in response to Ras activation. It is complicated to establish an order of

events given the data obtained. Overall MLL3 and CHD2 show the greatest effect on H3K4me3 while Aurora-B primarily impairs H3S10PK9Ac. In addition, knockdown of MLL3 and CHD2 seems to affect H3S10PK9Ac. Despite the clear effect on *Egr1* RNA synthesis and chromatin modifications, the knockdown of the modifiers selected seems to only modestly affect Pol II recruitment and escape. SMYD2 seemed specific for H3K4me3 at *Egr1* TSS potentially being specifically recruited by active Pol II at the transcriptional start site. Analysis of Pol II CTD modification and CDK9 recruitment are going to provide further insights into the transcriptional role of the chromatin modification studied.

6.5 Summary

In this chapter I presented preliminary data aimed at identifying chromatin modifiers and remodellers involved in TCF-dependent gene activation. A group of 7 hits was selected from a siRNA screen assessing the expression of *Egr1* and *Fos* at the pre-mRNA and mRNA level after knocking-down a panel of 50 target proteins. Aurora-B, MLL3, CHD2, Set7, Tip60, Tip49b and SMYD2 were selected. Only Aurora-B, MLL3, CHD2 and SMYD2 showed detectable impairment of defined marks at the *Egr1* promoter. Further analyses are going to be essential in order to corroborate these observed effects. Assessments of additional histone modifications are going to be pivotal in the understanding of the relationship between different chromatin marks.


A
Negative Controls
MOC
SCR
Positive Controls
MED23
SRF
Helicase
tip49a
tip49b
Remodelling
INO80
SWR1
SMARCA1
SMARCA2
smarcad1
CHD2
SMARCA4
CHD1
SMARCA5
Phosphatase
PPP1CB
PPP1CA
PPP1CC
PP2CB
PP2CA
Kinase
MSK1
IKK-a
MSK2
AURKA
AURKB
PIM1
RSK2
HAT
NCOA3
NCOA2
PCAF
NCOA1
ELP3
GCN5
p400
KAT3B
ESA1
MYST4
KAT5
MYST3
MOF
MYST2
KMT
MLL4
KMT3B
KMT2H
MLL3
MLL5
KMT3A
SETD1A
MLL2
KMT2G
KMT3C
MLL1
KMT3E
KMT7

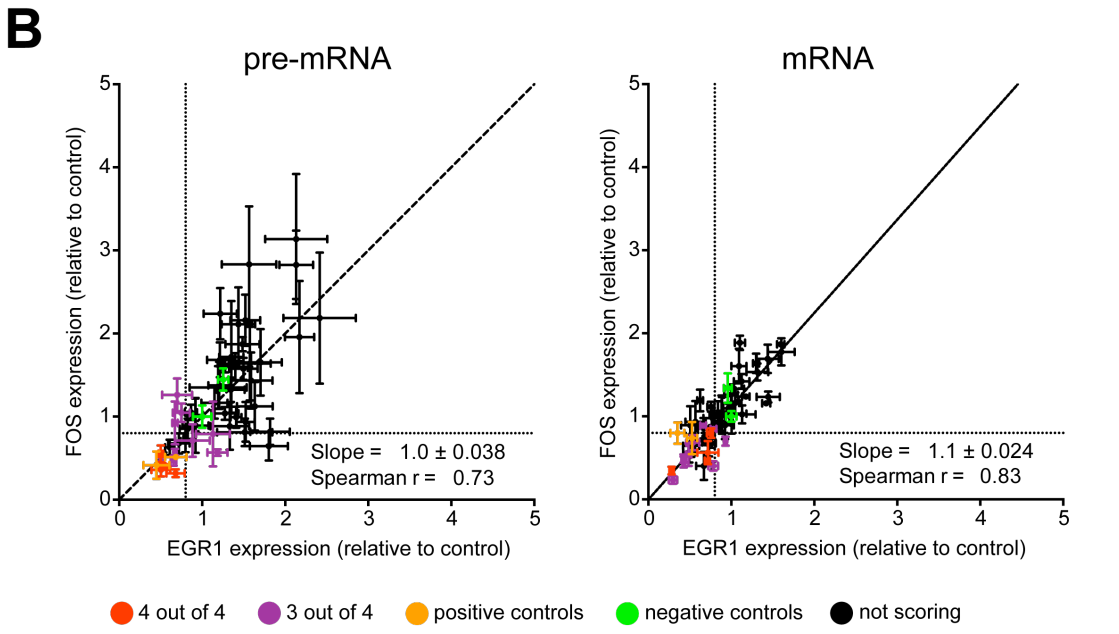

B
pre-mRNA
mRNA
FOS expression (relative to control)
EGR1 expression (relative to control)
Slope = 1.0 ± 0.038
Spearman r = 0.73
Slope = 1.1 ± 0.024
Spearman r = 0.83
4 out of 4
3 out of 4
positive controls
negative controls
not scoring

Figure 6.1 Phase 1 medium-throughput screening for chromatin-associated IE gene regulators in primary MEFs.

(A) Selected targets divided by activity. 50 siRNA against diverse chromatin modifiers and remodellers were selected including 5 phosphatases, 7 kinases, 9 remodellers, 14 acetyltransferases and 13 methyltransferases. Two negative controls including untreated cells (MOC) and a scrambled oligo (SCR) were used. Two positive controls including Med23 and SRF were included in the screen. **(B)** Scatter-plot representation of the results. (*Left*) Correlation between Fos and *Egr1* pre-mRNA, (*right*) correlation between Fos and *Egr1* mRNA. The values are relative to the negative controls (green). Non-scoring targets (black), 4 out of 4 (red), 3 out of 4 (purple) and the positive control Srf (yellow) are plotted. Dotted lines represent 20% reduction in expression.

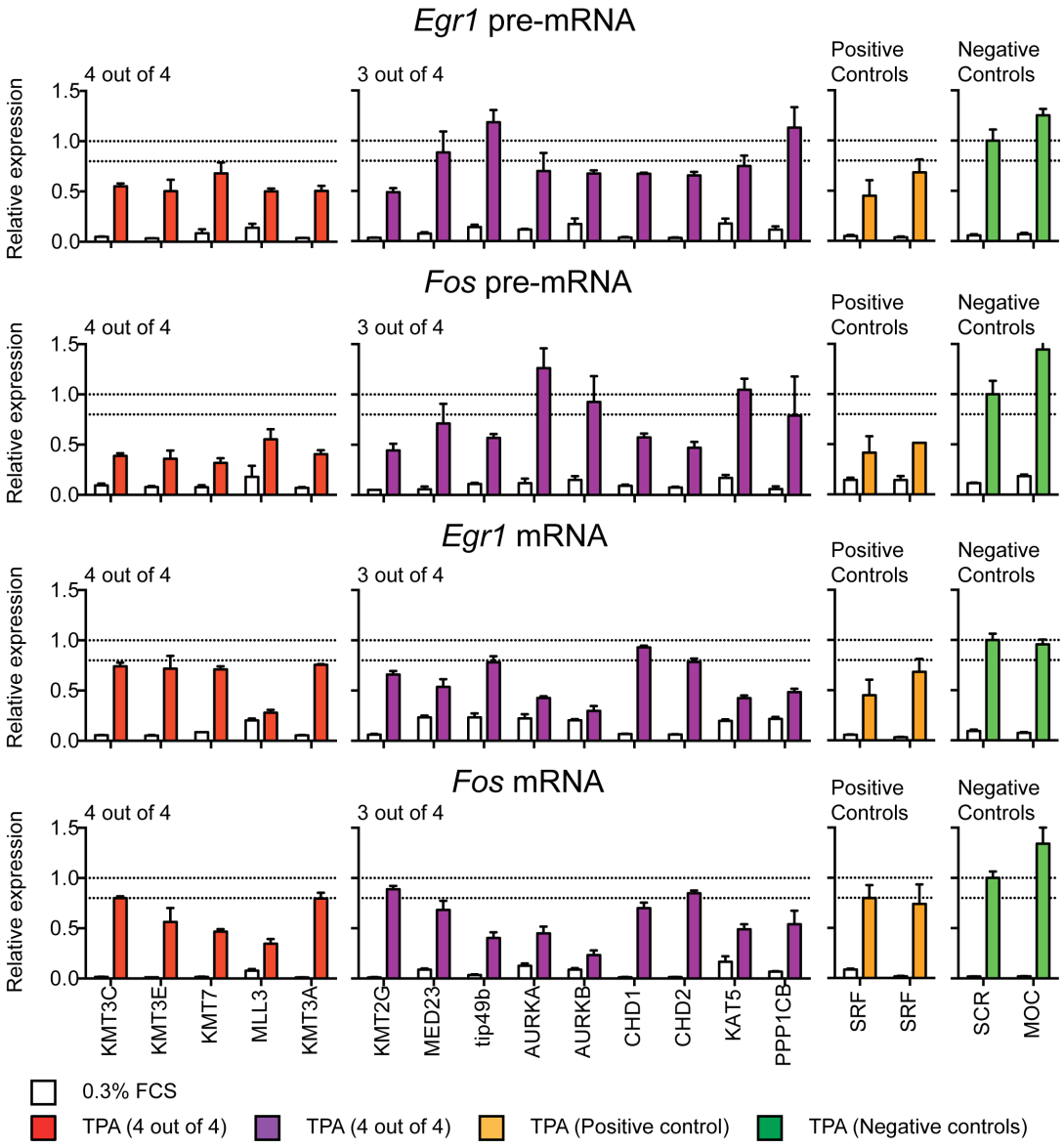


Figure 6.2 Phase one medium-throughput screen result.

Representation of the selected hits described in figure 6.1B. In each graph the top dotted line is 1 while the bottom line mark the 20% reduction threshold defined. Hits scoring in 4 out of 4 (red) and in 3 out of 4 (purple) readouts are shown. The knockdowns of SRF following TPA are shown in yellow. The transfections of scrambled sequence or untreated cells (MOC) are shown in green.

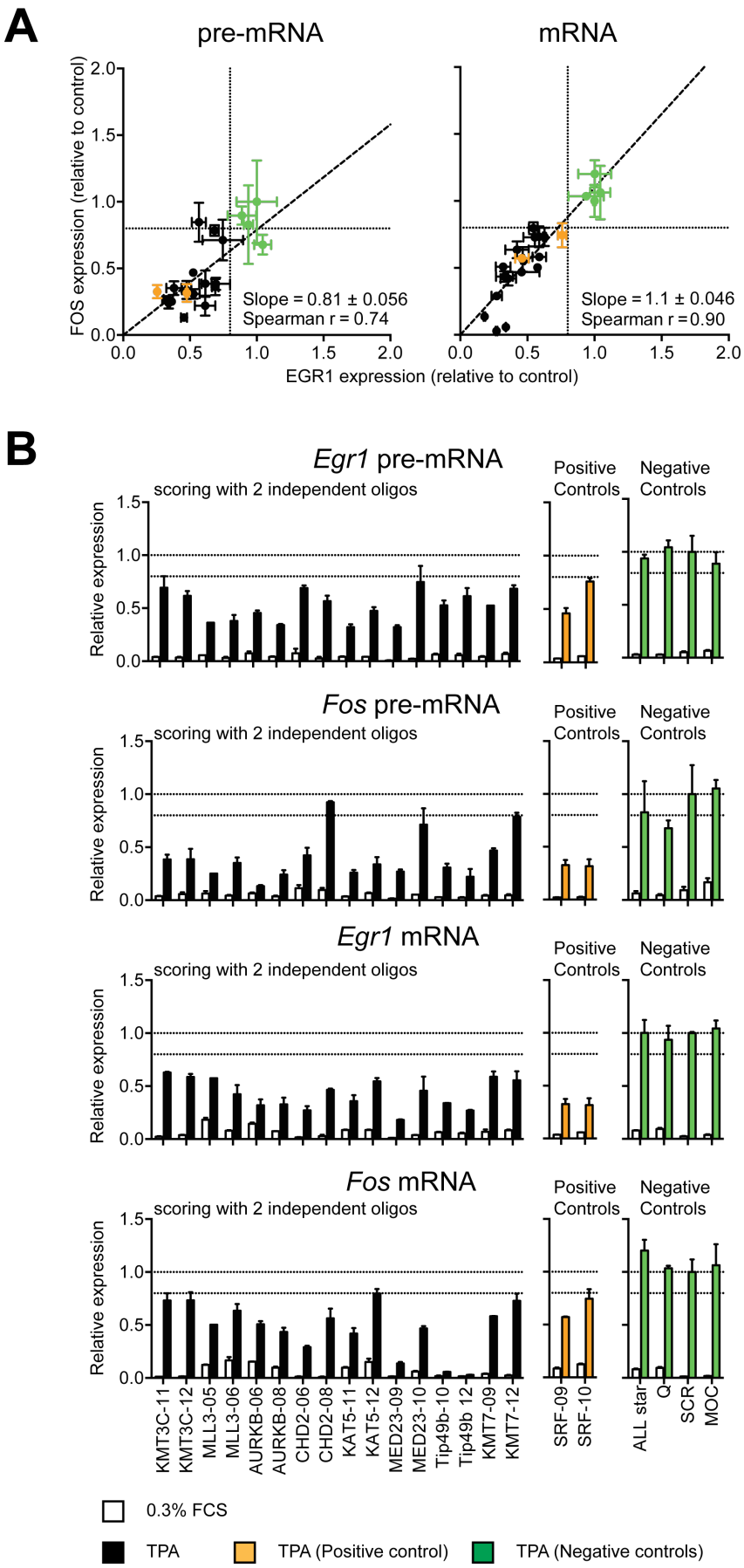

A
pre-mRNA
mRNA
FOS expression (relative to control)
EGR1 expression (relative to control)
Slope = 0.81 ± 0.056
Spearman r = 0.74
Slope = 1.1 ± 0.046
Spearman r = 0.90
B
Egr1 pre-mRNA
Fos pre-mRNA
Egr1 mRNA
Fos mRNA
scoring with 2 independent oligos
Positive Controls
Negative Controls
Relative expression
KMT3C-11
KMT3C-12
MLL3-05
MLL3-06
AURKB-06
AURKB-08
CHD2-06
CHD2-08
KAT5-11
KAT5-12
MED23-09
MED23-10
Tip49b-10
Tip49b 12
KMT7-09
KMT7-12
SRF-09
SRF-10
ALL star
Q
SCR
MOC
0.3% FCS
TPA
TPA (Positive control)
TPA (Negative controls)

Figure 6.3 Phase 2 medium-throughput screening for chromatin-associated IE gene regulators in primary MEFs.

(A and B) Deconvolution experiment showing in black the 7 hits impairing both Fos and *Egr1* pre-mRNA and/or mRNA. The results obtained for 22 oligos are displayed: 2 for each selected hit (black), 2 for the Med23 positive control (black), 2 for the Srf positive control (yellow) and 4 negative controls (green).

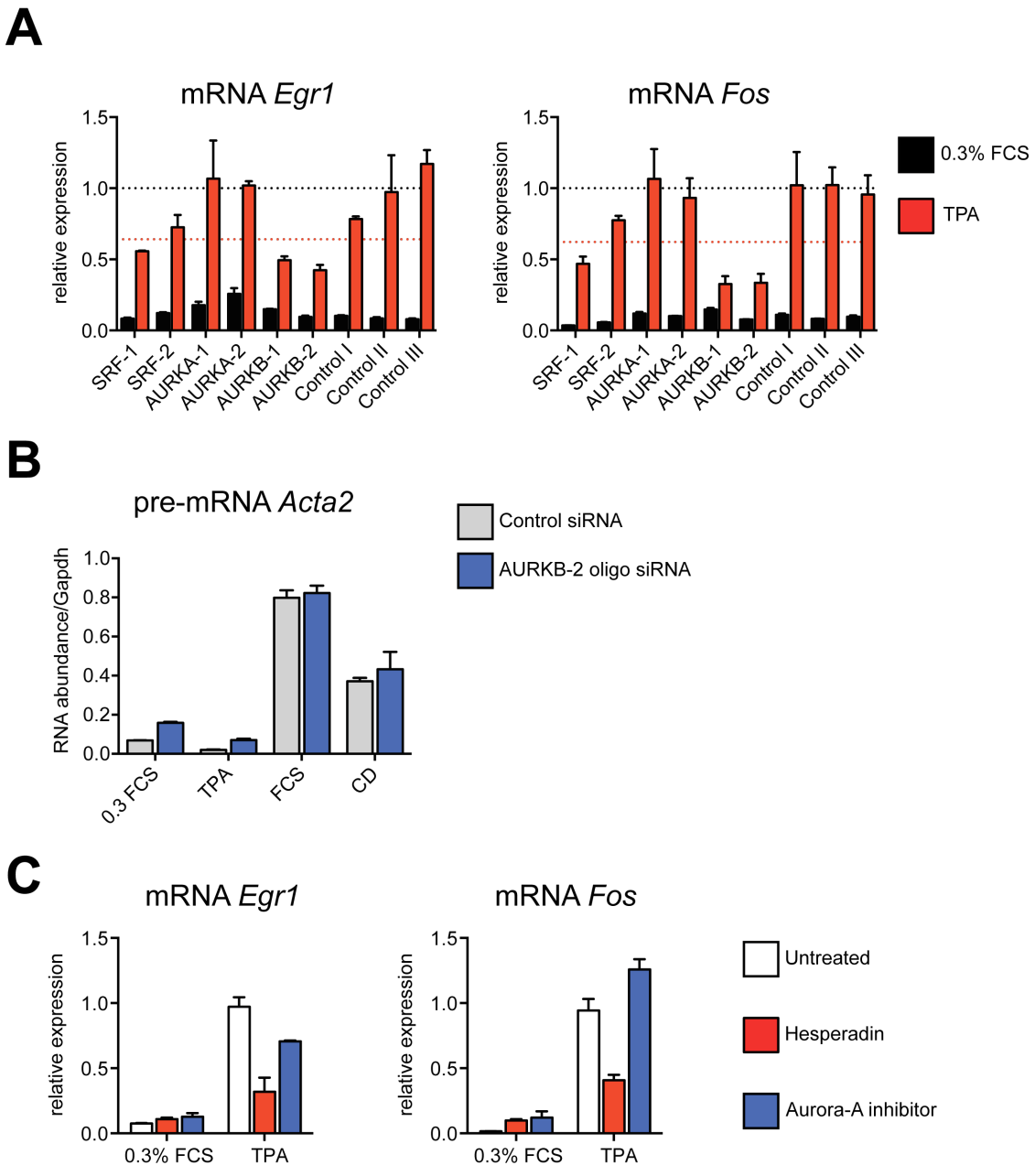


Figure 6.4 Aurora-B validation

(A) Comparison of *Egr1* and Fos expression following Srf, Aurora-A and Aurora-B knockdown using two independent oligos for each target. **(B)** Effect of Aurora-B knockdown on Acta2 expression following TPA, FCS or CD. **(C)** Aurora-B (Hesperadin, 200nM) and Aurora-A specific drug (Aurora Kinase Inhibitor III at 200nM, Calbiochem) treatments assessing *Egr1* and *Fos* expression. 5 minutes pre-treatment step with the indicated compounds followed by 30 minutes TPA stimulation.

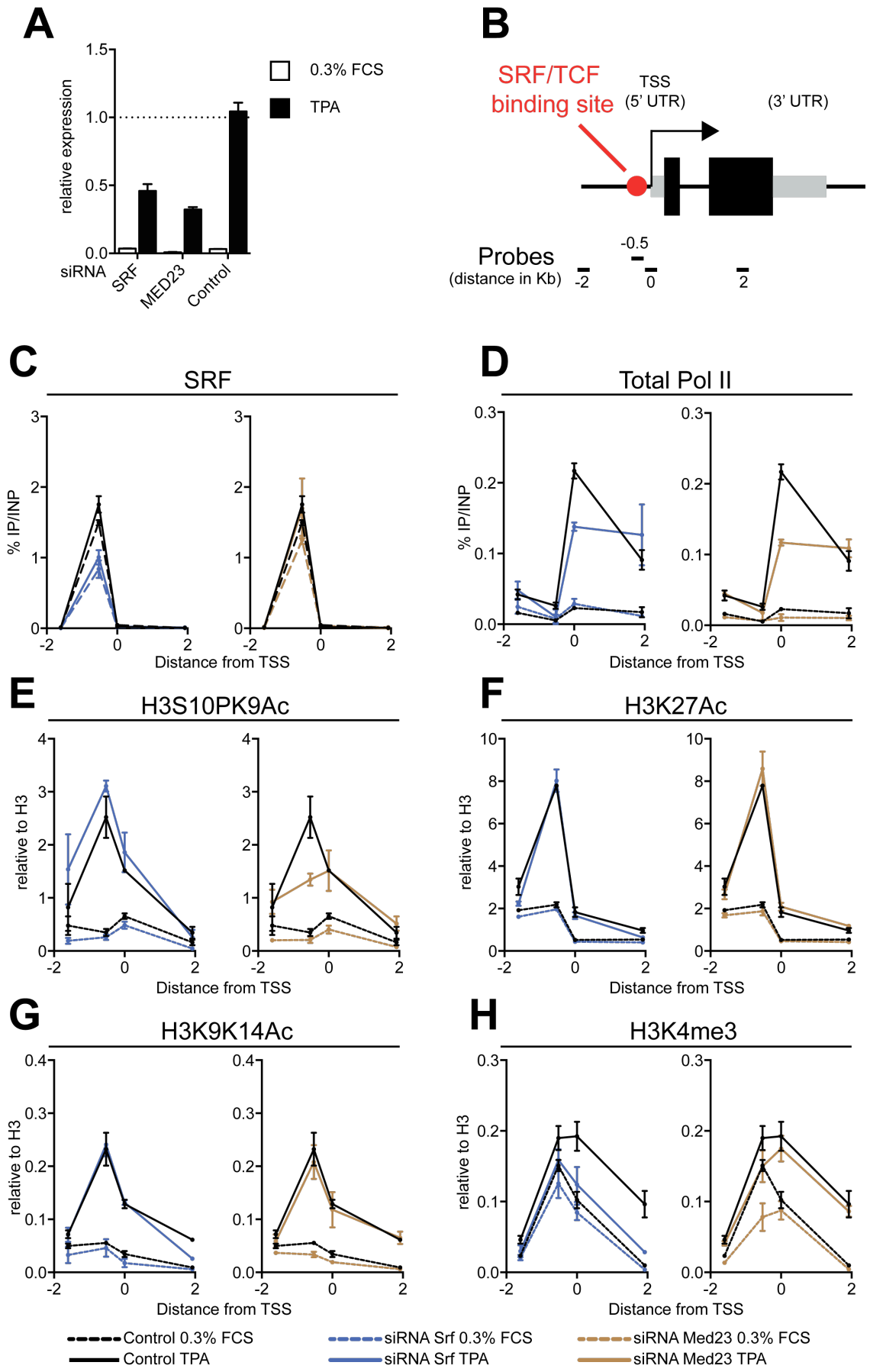

A
relative expression
1.5
1.0
0.5
0.0
0.3% FCS
TPA
siRNA
SRF
MED23
Control
B
SRF/TCF binding site
TSS (5' UTR)
(3' UTR)
Probes (distance in Kb)
-0.5
-2
0
2
C
SRF
% IP/INP
Distance from TSS
D
Total Pol II
% IP/INP
Distance from TSS
E
H3S10PK9Ac
relative to H3
Distance from TSS
F
H3K27Ac
relative to H3
Distance from TSS
G
H3K9K14Ac
relative to H3
Distance from TSS
H
H3K4me3
relative to H3
Distance from TSS
Control 0.3% FCS
Control TPA
siRNA Srf 0.3% FCS
siRNA Srf TPA
siRNA Med23 0.3% FCS
siRNA Med23 TPA

Figure 6.5 siRNA ChIP experiment setup following SRF and Med23 knockdown.

(A) RT-PCR of *Egr1* pre-mRNA after TPA stimulation following knockdown of either SRF or Med23. **(B)** Representation of the *Egr1* gene illustrating the positions of different probes along the gene. Exons are indicated with 5'UTR and 3'UTR in grey and coding sequences in black; SRF binding site is indicated with a red circle. **(C to H)** ChIP experiment, following transfection of cells with siRNA against SRF (blue), Med23 (brown) or Scramble (black), with the indicated antibody across *Egr1* promoter and ORF as in Chapter 5.

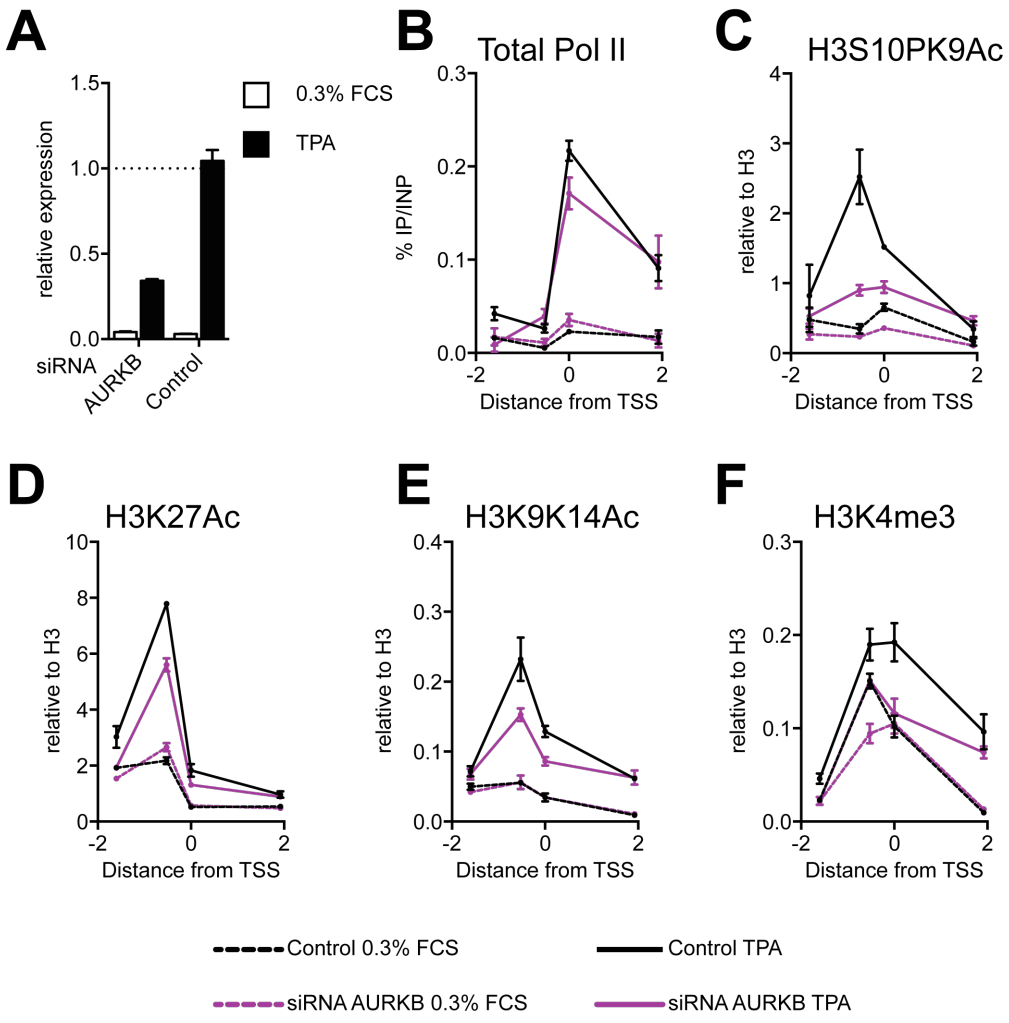


Figure 6.6 Aurora-B dependent histone modifications in response to Ras activation.

(A) RT-PCR of *Egr1* pre-mRNA after TPA stimulation following knockdown of AuroraB. **(B to G)** ChIP experiment, following transfection of cells with siRNA against AuroraB (purple) or Scramble (black), with the indicated antibody across *Egr1* promoter and ORF as in Chapter 5.

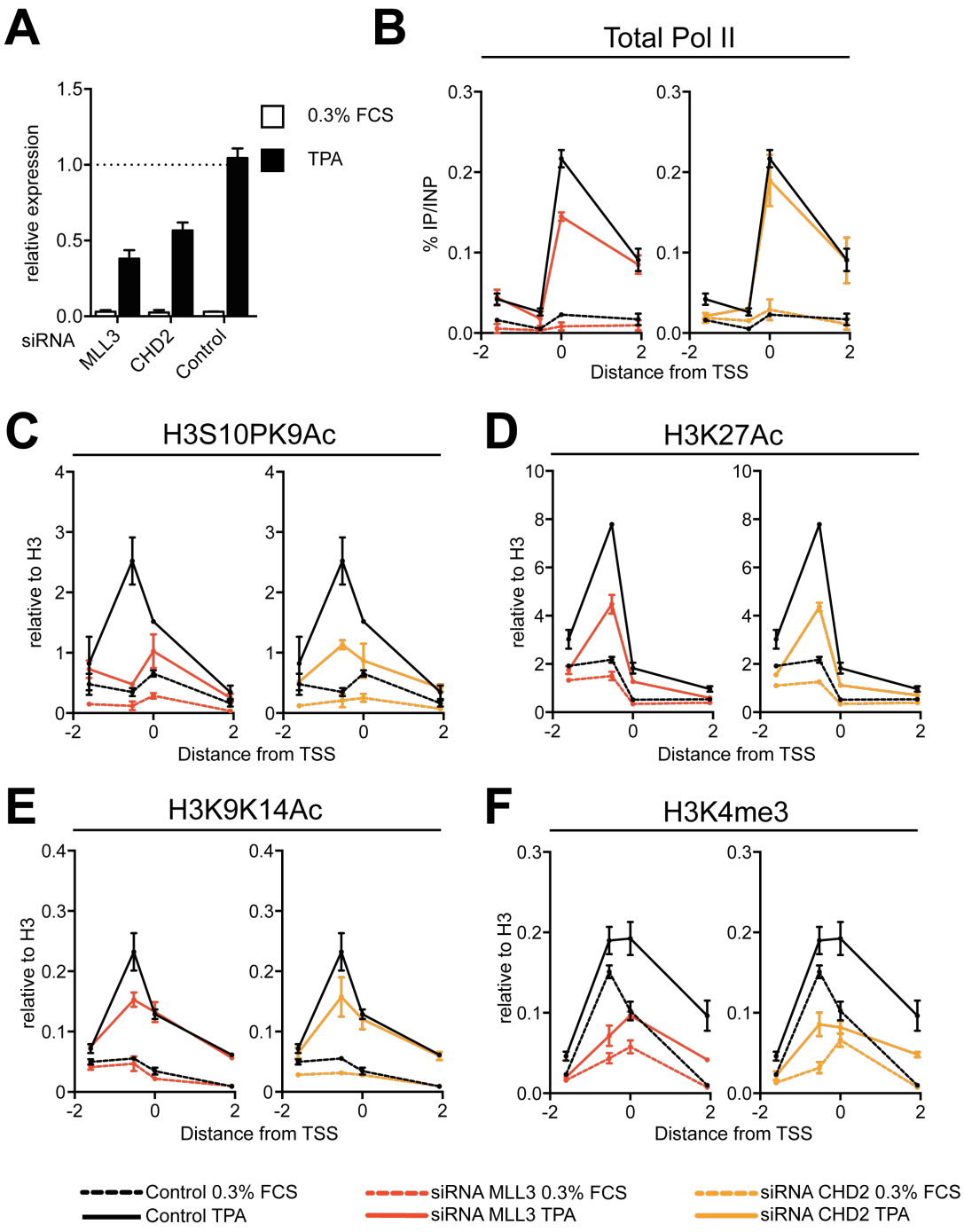


Figure 6.7 MLL3 and CHD2 dependent histone modifications in response to Ras activation.

(A) RT-PCR of *Egr1* pre-mRNA after TPA stimulation following knockdown of MLL3 and CHD2. **(B to F)** ChIP experiment, following transfection of cells with siRNA against MLL3 (red), CHD2 (yellow) or Scramble (black), with the indicated antibody across *Egr1* promoter and ORF as in Chapter 5.

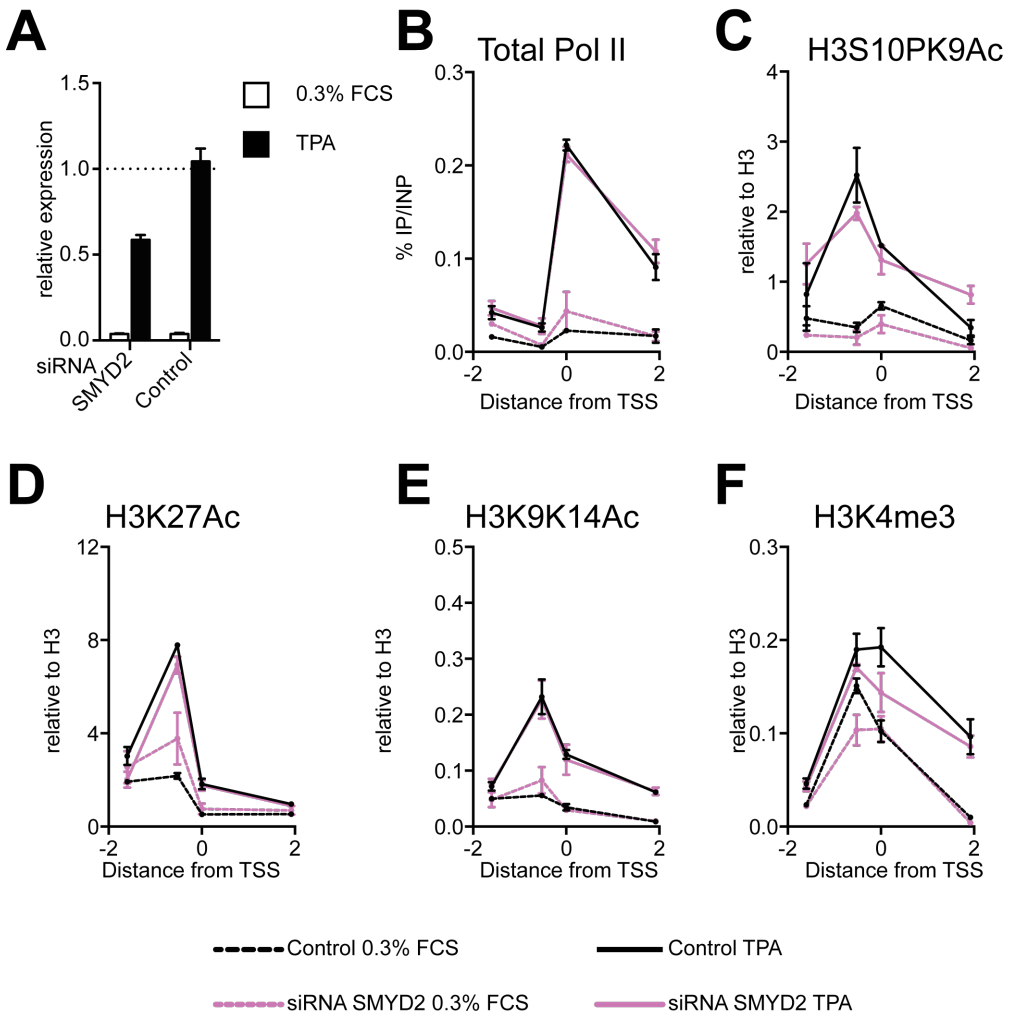


Figure 6.8 SMYD2 dependent histone modifications in response to Ras activation.

(A) RT-PCR of *Egr1* pre-mRNA after TPA stimulation following knockdown of SMYD2. **(B to G)** ChIP experiment, following transfection of cells with siRNA against SMYD2 (purple) or Scramble (black), with the indicated antibody across *Egr1* promoter and ORF as in Chapter 5.

Chapter 7. A genomic perspective on TCF-dependent gene activation

7.1 Aim

In Chapters 5 and 6, based on *bona fide* TCF target genes, I described the pivotal role of the TCFs in coordinating chromatin changes and transcriptional activation in response to Ras activation. In this chapter I am going to present preliminary data aiming to generalise these observations using genomic approaches. Initially I am going to dissect the response to TPA stimulation in wild type and TCFs knock out MEF cell lines using RNA-seq approaches. Focus will be dedicated primarily to TPA induced genes and genes that don't respond to TPA but whose baseline expression required the TCFs. TPA-repressed and TCF-repressed genes are also going to be considered. This analysis will be crossed with RNA-seq data obtained assessing the TPA response in reconstituted cell lines described earlier in Chapter 5. A functional analysis using DAVID is going to highlight the role of the TCF in cell growth and proliferation. Finally I am going to assess changes of a few chromatin marks genome-wide in different cell background for a selected set of Elk-1 controlled genes.

7.2 Genomic dissection of TCF-controlled genes

To determine the transcriptional output following TPA stimulation we initially analysed RNA-seq data obtained from wild type and ESN MEFs. As mentioned in Chapter 5 the cell lines used are phenotypically different. In particular it was possible to observe a reduced growth for cells lacking all three TCFs with a corresponding blockage in the cell cycle transition (Figure 7.1). Reconstitution with Elk-1 protein did not restore normal growth suggesting the possibility that either Sap1, Net or a combination of all three TCFs is required in order to maintain the expression of genes involved in cell cycle and proliferation (Cyril Esnaut personal communication). These observations are going to be essential for the interpretation of the ontology analysis that I will present later in this Chapter, as several indirect effects could also affect our analysis.

A differential gene expression analysis, based on a negative binomial distribution (Deseq), was applied. 2898 genes were shown to be induced while 2130 showed a reduced expression in wild type MEF following TPA stimulation (Figure 7.2A). 871 genes, induced by TPA in wild type MEF, were shown to be impaired in ESN MEF. On the other hand 531 genes, whose expression was reduced upon TPA stimulation, were shown to have a higher expression in ESN MEF in both resting and TPA stimulated conditions (Figure 7.2 B). Although statistically significant the differences between unstimulated and TPA repressed are subtle. Furthermore it is possible to observe that this group of genes shows a higher baseline in ESN MEF if compared to wild type MEF, possibly caused by a TCF-dependent repression of their expression (Figure 7.2 B).

We also considered genes non-induced by TPA whose baseline was changing in ESN MEF if compared to wild type MEF. 3745 genes showed a lower expression in ESN MEF suggesting a possible role of the TCFs in positively controlling their baseline expression (Figure 7.2 C). On the other hand 3627 genes showed a higher expression in ESN MEF if compared to wild type MEF (Figure 7.2C).

As strictly relevant to the project we initially focused on TPA-induced genes. Considering the behaviour of the model genes presented earlier, we can consider three transcriptional signatures for genes induced by TPA. Genes like *Egr1*, besides being less induced by TPA in ESN MEF, are showing a reduced baseline expression. This transcriptional signature could be assigned to genes with a transcriptional outcome that stringently relies on the TCFs. Genome-wide we could select 507 genes out of the 871 set with an impaired baseline in ESN MEF (Figure 7.3A). A second class including *Egr2* and *Fosl1*, composed of 320 genes, showed no change in baseline expression, probably due to the activity of further ETS members controlling their expression (Figure 7.3A). Finally a small group of 44 genes, including *Nr4a1*, *Fos* and *Ier2*, showed a higher baseline in ESN MEF although their induced expression upon TPA was affected in knockout cell (Figure 7.3A). Investigation of specific sequence motifs differentially enriched at the promoters of genes within each group will be highly informative to further elucidate the transcriptional signatures observed. The analysis proceeded including the RNA-seq data set obtained with cells reconstituted with Elk-1 wild type, FW and Nona. These rescue experiments allowed us to identify 92 genes where TPA

induced expression could be restored with Elk-1 wild type (Figure 7.3 B). On the other hand reconstitution with FW or Nona did not, suggesting that the selected targets are indeed Elk-1 controlled (Figure 7.3B). The selected group of genes looked fairly small if compared to the 792 genes impaired in ESN MEF. Despite the functional redundancy that has been reported between Elk-1 and SAP-1 (Costello et al. 2010), is conceivable that at the genomic scale differences could be observed. Sap-1 or even Net could potentially control the remaining genes where Elk-1 seems insufficient.

As mentioned above we conducted the same analysis on TPA repressed genes. As mentioned earlier this group of genes showed a higher baseline expression in ESN MEF if compared to wild type MEF (435 out of 531) (Figure 7.4 A). Only 92 showed no change in baseline expression and 4 genes showed a reduced baseline in ESN MEF (Figure 7.4 A). The changes observed across conditions for the last two groups were subtle and possibly excluded by increasing the stringency of the statistical test used. Reconstitution with Elk-1 wild type was shown to re-establish a statistical significant reduction following TPA stimulation at just 16 of the selected genes (Figure 7.4 B). Also cells reconstituted with Nona and FW showed a lower expression of these genes following TPA (Figure 7.4 B). This observation suggests that the repression observed is not directly caused by the Elk-1 AD.

Considering genes whose baseline expression was changing in ESN MEF, a wider group of 507 genes was rescued by the ectopic expression of Elk-1 wild type if compared to the control cell line (Figure 7.5 A). Consistent with a potential function of uniquely controlling the baseline expression of these genes, ESN reconstituted with Elk-1 Nona or FW showed a higher baseline if compared to the control (Figure 7.5 B). This observation suggests that these genes only require Elk-1 DNA binding activity rather than its ability to respond to Ras signalling. On the other hand 587 genes, whose expression was enhanced in ESN MEF, showed a reduced expression when Elk-1 was ectopically expressed (Figure 7.5 C). Also for these groups of genes the Elk-1 mutants Nona and FW were lowering the baseline if compared to ESN infected with vehicle (Figure 7.5 D).

In conclusion the TCFs control 30% of the TPA response in MEF cells. Elk-1 is sufficient to control 10% of the TCF-dependent genes suggesting that Sap-1 or Net could probably contribute to the control of the remaining TCF-dependent set.

Further analysis including cell reconstitution with the other TCF members and a description of the different promoter context is going to provide further insights into the TCF-dependent gene regulation.

7.3 TCFs regulate cell proliferation and cell cycle progression

We analysed the functional categories enriched in each selected group of genes using DAVID (Huang et al. 2009). Genes potentially controlled by the TCFs were revealed to be involved mainly in transcriptional regulation and cell cycle control (Figure 7.6). A similar signature was also found when we analysed the 92 genes directly controlled by Elk-1 (Figure 7.6). Furthermore genes controlled at the baseline level by Elk-1 showed almost identical functional categories. The ontology analysis was also performed for repressed genes. TCFs seemed involved in the repression of genes involved in cell differentiation, tissues and organ development and actin-linked processes. ESN MEF reconstituted with Elk-1 seemed to also negatively control some genes with similar functions.

The identification of genes involved in cell cycle progression was intriguing as it has been recently described that Elk-1 together with Erk2 could control cell-cycle related genes (Göke et al. 2013). As introduced earlier within this chapter ESN MEF are blocked in the cell cycle showing a reduced proliferation. Despite a clear enrichment in cell cycle related functions for genes controlled by Elk-1 it was not possible to observe the reestablishment of a correct proliferation for ESN MEF reconstituted with Elk-1 wild type (see section 7.2). This observation implies that SAP-1 and possibly Net are also required for a correct cell cycle progression. On the other hand this observation also implies that the functional categories enriched for the Elk-1 specific set are not indirectly caused by the reestablishment of proliferation. Elk-1 seems therefore to directly control genes involved in cell cycle, rather than controlling a few pro-proliferative genes that could promote a secondary response in cell cycle related genes.

Reconstitution with the other TCF members is going to be pivotal in understanding the potential role of the TCFs in controlling cell cycle progression. It is clear from the analysis conducted that Elk-1, although incapable of restoring the complete transcriptional response in ESN MEF, is still able to control a considerable fraction of cell cycle related genes.

7.4 TCF-dependent shaping of chromatin modification at target promoters

The analysis conducted above, besides describing the transcriptional signature in response to Ras activation, had the aim of selecting a reliable set of Elk-1 targets with which we could generalise the observations listed in Chapter 5. We combined the ChIP experiments described in Chapter 5 with deep sequencing. Here I will briefly describe the profiles observed for H3K4me3 and H3K9K14ac (Figure 7.8). The data presented does not take into account changes in histone occupancy. We are currently analysing changes in distribution of the various chromatin modifications specifically at the SRF/TCF binding sites in order to gain further insights and corroborate the data collected based on the Egr1 model gene.

We conducted a preliminary analysis focusing on changes observed at the TSS of targets whose expression was shown to be Elk-1 dependent and sensitive to both FW and Nona mutations at p-values less than 0.05 (Figure 7.7). The selected group of targets showed to be overall closer to SRF. Both H3K4me3 and H3K9K14ac in a wild type context were shown to be enhanced upon TPA stimulation and to invade the area downstream of the TSS (Figure 7.9 A and 7.10 A). ESN MEF did not show such an enhancement for either modification. Reconstitution experiments using Elk-1 wild type re-established the profiles observed in wild type MEF for both modifications (Figure 7.9 B and 7.10 B). The situation observed for ESN MEF rescued with Elk-1 FW or Nona was shown to be different. While Elk-1 Nona showed no induction of either modification (Figure 7.9 D and 7.10 D), Elk-1 FW allowed a partial enhancement of both H3K9K14ac and H3K4me3 close to the TSS (Figure 7.9 C and 7.10 C). The data presented does not consider changes in histone occupancy around the TSS potentially affecting the signal distribution of the observed histone modifications.

Despite being only preliminary, these data confirm that chromatin changes can occur in the absence of transcription. A detailed analysis of the spatial distribution of each chromatin modification together with changes in histone occupancy is going to further elucidate the relationship between transcriptional activation and chromatin modifications. Furthermore it is going to be crucial to analyse changes observed at the SRF-TCF binding sites. In particular is going to be important to analyse the distribution of H3K4me3 at the SRF/TCF binding site in

TCF null fibroblasts reconstituted with Elk-1 Nona and FW in order to confirm the observations collected using the *Egr1* model gene.

7.5 Summary

In this chapter, I have presented our initial approach for the generalisation of the observations collected and presented in Chapter 5 and 6. Using an RNA-seq based approach it was possible to describe the transcriptional response following TPA stimulation in cells lacking all TCFs. In particular TCFs control 30% of the genes activated in response to TPA. Reconstitution experiments showing that Elk-1 suffices for the expression of 10% of these genes imply that SAP-1 and Net are also required to control the rest of the selected targets. Analysis of H3K4me3 and H3K9K14ac across the different cell lines showed that signal-induced chromatin changes could be disentangled from the transcriptional process. This analysis is going to be extended to the other chromatin modifications presented in Chapter 5 considering the diverse groups of genes described with the RNA-seq analysis. Furthermore we will also consider changes at the SRF-TCF binding sites across the different reconstituted cell lines.

The RNA-seq analysis was also extended considering genes whose expression was unaffected by TPA but whose baseline activity was controlled by the TCFs. 20% of these genes were rescued by Elk-1 ectopic expression. Genes positively controlled by the TCFs were functionally involved in transcription and cell cycle control. On the other hand genes whose expression was impaired by the TCFs were involved in cell differentiation and development. Indeed, cells lacking all TCFs showed a reduced proliferation if compared to wild type MEF. Furthermore reconstitution experiments using Elk-1 wild type showed no change in proliferation rate while the rescued genes could be assigned to the same functional classes. This observation allowed us to exclude indirect effects possibly caused by re-establishing the proliferation. These data are going to be corroborated by reconstituting cells with SAP-1, Net or a combination of the three TCFs.

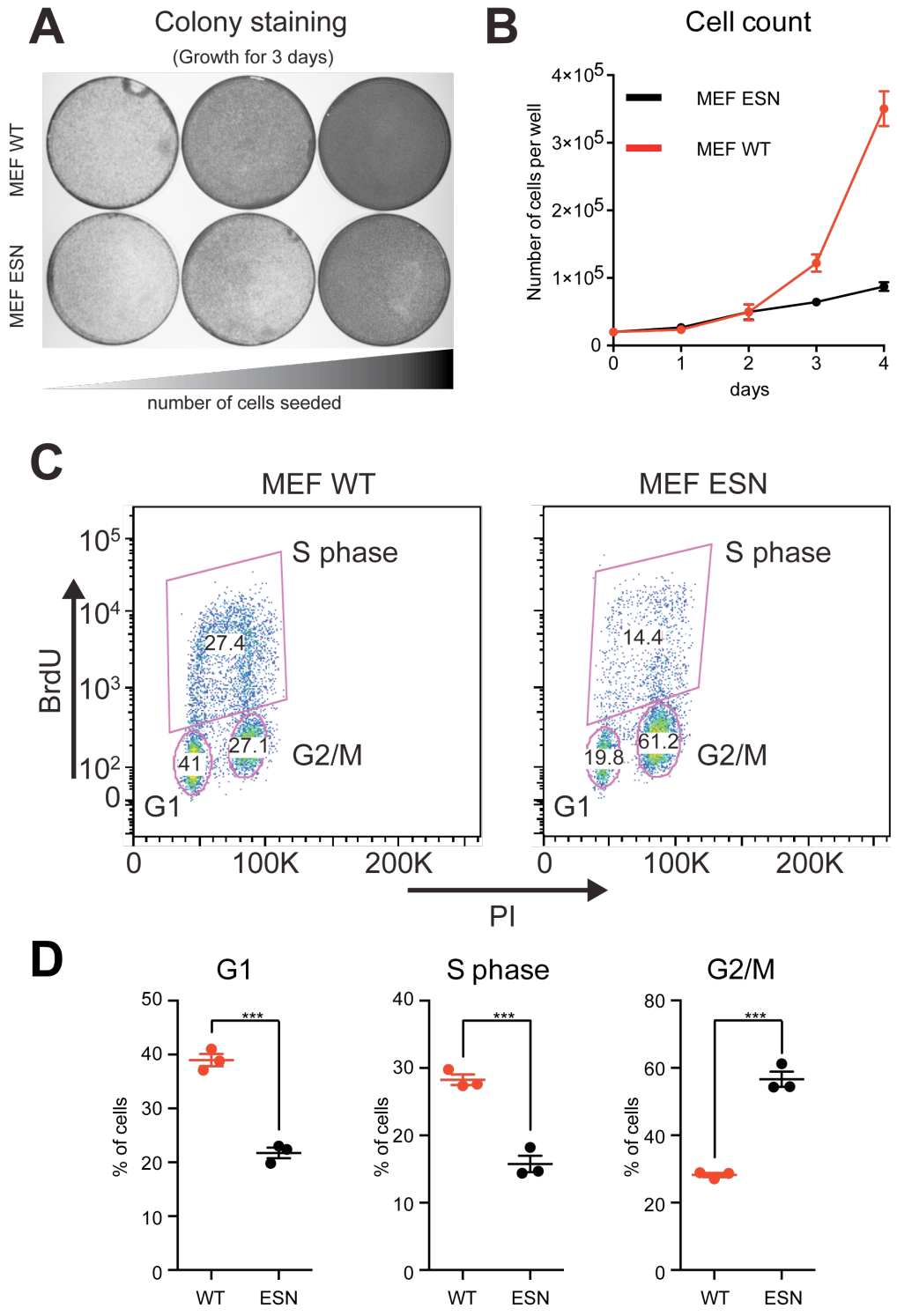


Figure 7.1 TCFs regulate MEF proliferation

(A) Crystal violet staining of cells grown for three days at different seeding density. (*Top*) wild type MEF (MEF WT) (*Bottom*) ESN MEF. **(B)** Cell proliferation curve over 4 days. Cells were harvested and counted using flow cytometry. (red) wild type MEF, (black) ESN MEF. **(C)** Analysis of the DNA content using propidium Iodate (PI) staining against Bromodeoxyuridine (BrdU) incorporation performed for 4h. **(D)** Quantification of each cell cycle phase done in triplicate as in panel C.

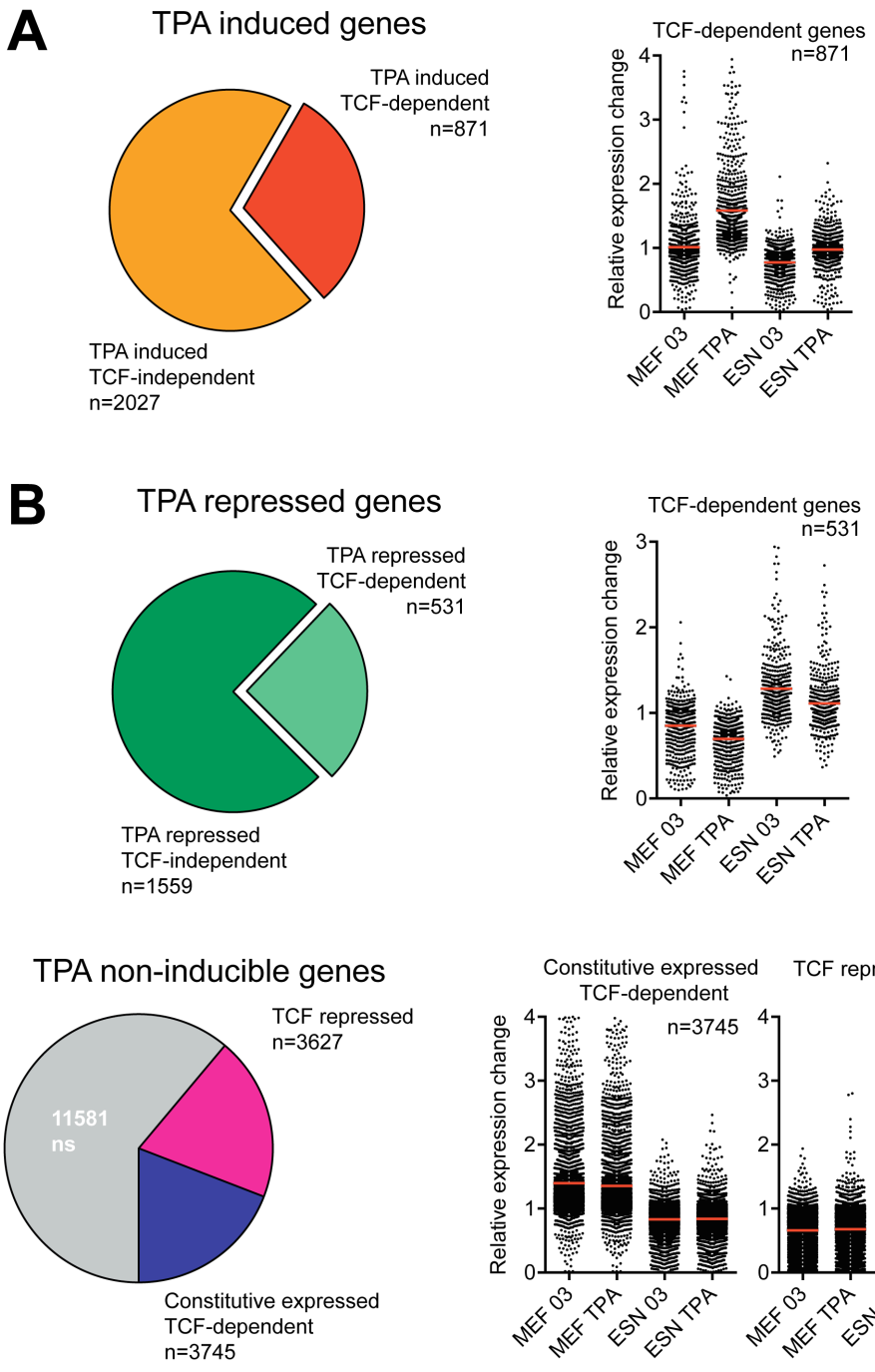

A
TPA induced genes
TPA induced
TCF-dependent
n=871
TPA induced
TCF-independent
n=2027
TCF-dependent genes
n=871
Relative expression change
MEF 03
MEF TPA
ESN 03
ESN TPA
B
TPA repressed genes
TPA repressed
TCF-dependent
n=531
TPA repressed
TCF-independent
n=1559
TCF-dependent genes
n=531
C
TPA non-inducible genes
TCF repressed
n=3627
11581
ns
Constitutive expressed
TCF-dependent
n=3745
Constitutive expressed
TCF-dependent
n=3745
TCF repressed
n=3627

Figure 7.2 TCF-dependent gene control in MEF

RNA-seq analysis of genes induced or repressed by TPA (panel A and B) or whose baseline expression changes in ESN MEF (panel C). **(A)** Genes induced by TPA in wild type MEF are shown. (*Left*) 871 TPA-induced genes are affected in ESN MEF (red quadrant), being probable TCF-dependent targets, while 2027 are still induced in ESN MEF upon TPA stimulation (yellow quadrant). (*Right*) relative expression of the TCF-dependent TPA induced genes (n=871) in wild type and ESN MEFs before and after TPA stimulation. **(B)** Genes repressed by TPA in wild type MEF are shown. (Left) 531 TPA-repressed genes are affected in ESN MEF (light green quadrant) while 1559 are still repressed in ESN MEF upon TPA stimulation (dark green quadrant). (Right) relative expression of the TCF-dependent TPA-repressed genes (n=531) in wild type and ESN MEFs before and after TPA stimulation. Genes repressed upon TPA in wild type MEF were analysed in ESN MEF. We considered as TCF dependent those genes that were not anymore showing a statistical significant repression in ESN MEF following TPA **(C)** Genes non-inducible by TPA are shown. (Left) 11581 show no change comparing wild type and ESN MEFs. 3745 requires the TCFs for their basal expression (blue) while 3627 are active in the absence of the TCF (pink). (Right) Relative expression for TCF-induced and repressed in wild type and ESN MEFs.

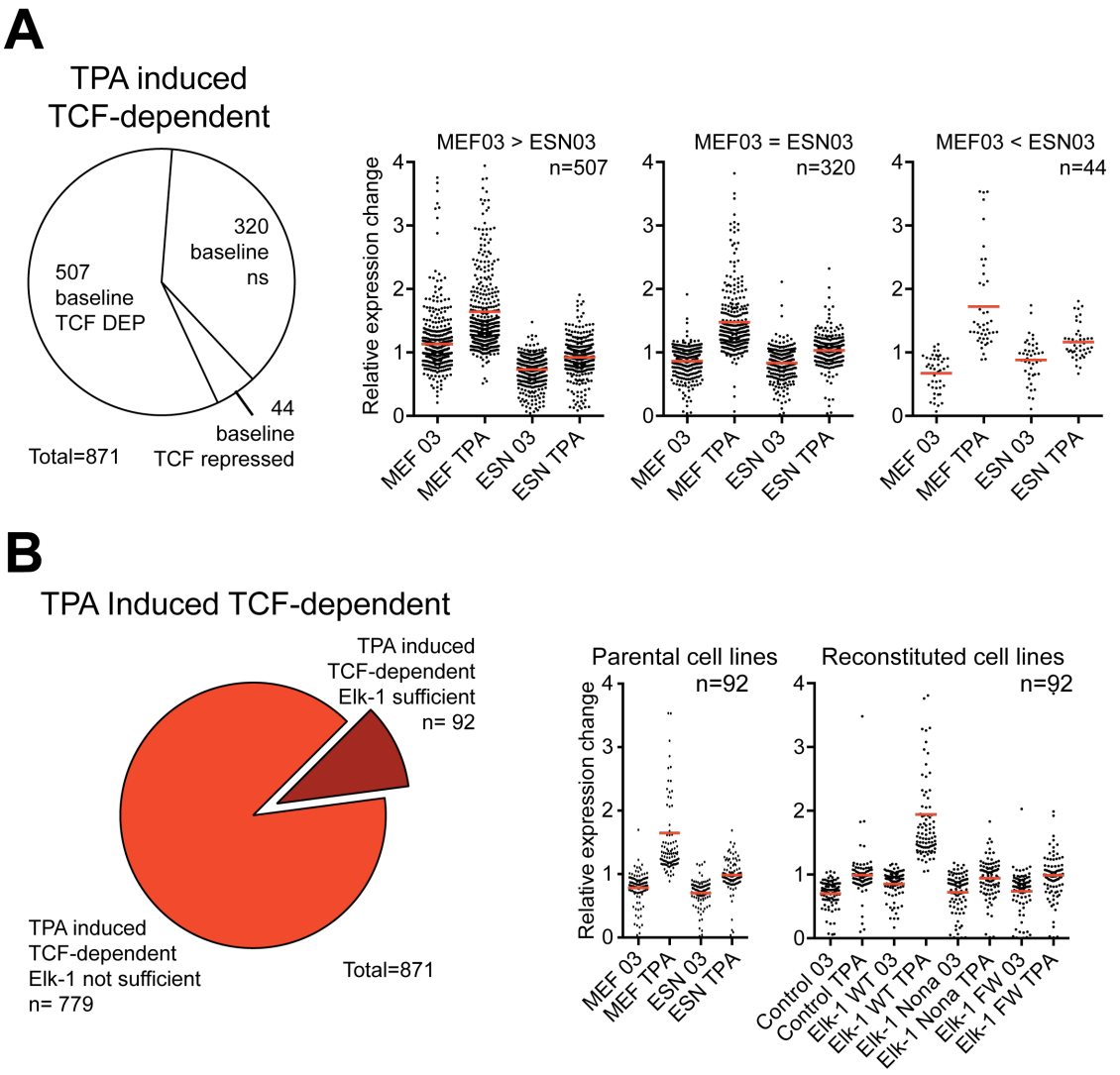


Figure 7.3 Elk-1 controls 10% of the TCF-dependent TPA-induced targets

(A) (Left) Analysis of the baseline expression in ESN MEF against wild type MEF for TPA-induced TCF-dependent genes (n=871). 507 genes induced by TPA in a TCF-dependent manner showed reduced baseline in ESN MEF. 320 showed no changes in baseline expression and 44 showed enhanced baseline in ESN MEF. (Right) Relative expression across cell lines and conditions for the selected group of genes described**. (B)** Elk-1 is sufficient to control 10% of the TPA-induced TCF-dependent genes. (Left) 92 targets out of the 871 selected showed to be re-activated in ESN MEF reconstituted with Elk-1 wild type. (Right) Relative expression across cell lines for the 92 selected targets. Parental lines uninfected and reconstituted cell lines are displayed.

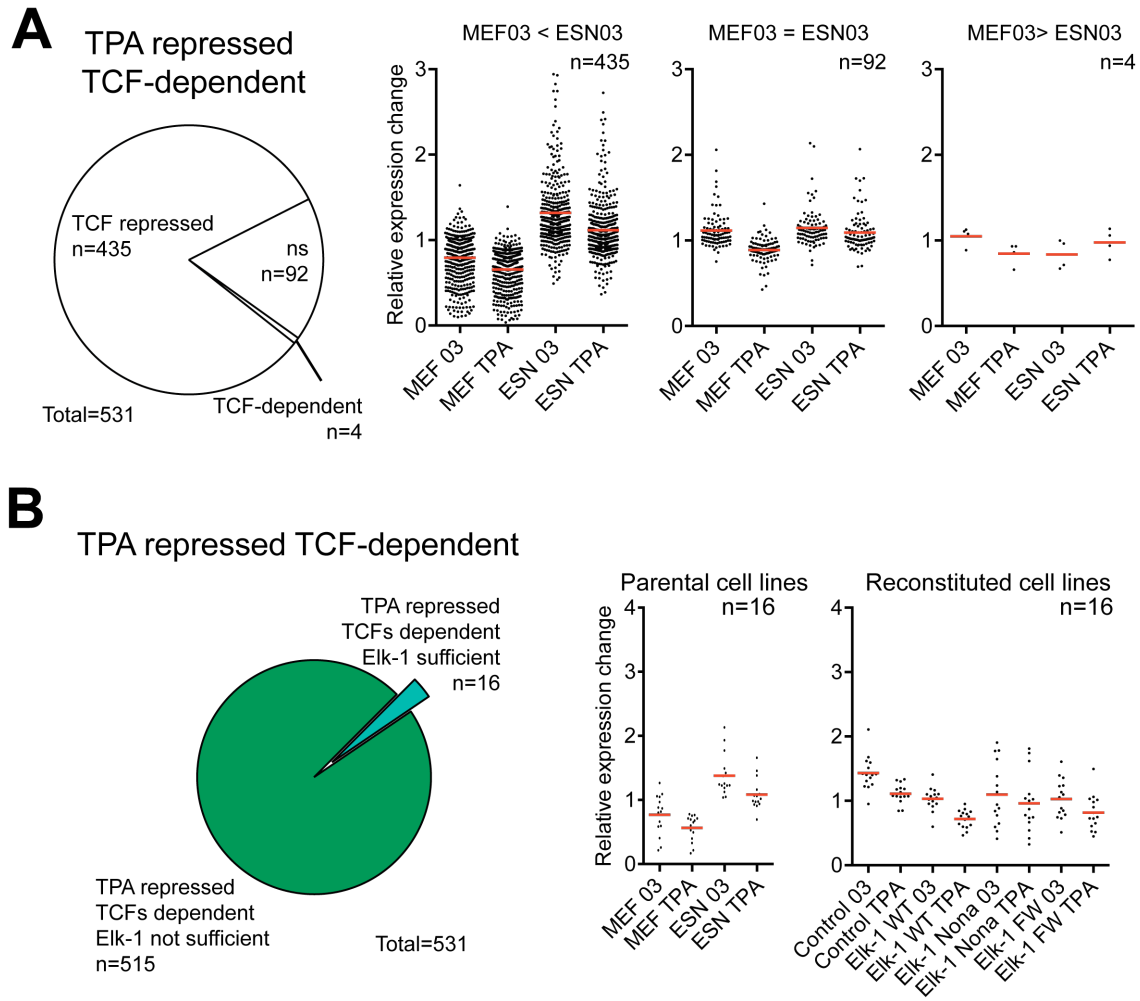


Figure 7.4 TCFs repress the baseline expression of defined targets

(A) (Left) Analysis of the baseline expression in ESN MEF against wild type MEF for TPA-repressed TCF-dependent genes (n=531). 435 genes repressed by TPA in a TCF-dependent manner showed enhanced baseline in ESN MEF. 92 showed no changes in baseline expression and 4 showed reduced baseline in ESN MEF. (Right) Relative expression across cell lines and conditions for the selected group of genes described. **(B)** Reconstitution with Elk-1 wild type is able to re-establish the TPA-dependent repression only at 16 targets. (Left) Pie chart displaying the 531 TPA repressed TCF-dependent targets. (Right) Relative expression across cell lines for the 16 selected targets. Parental lines uninfected and reconstituted cell lines are displayed.

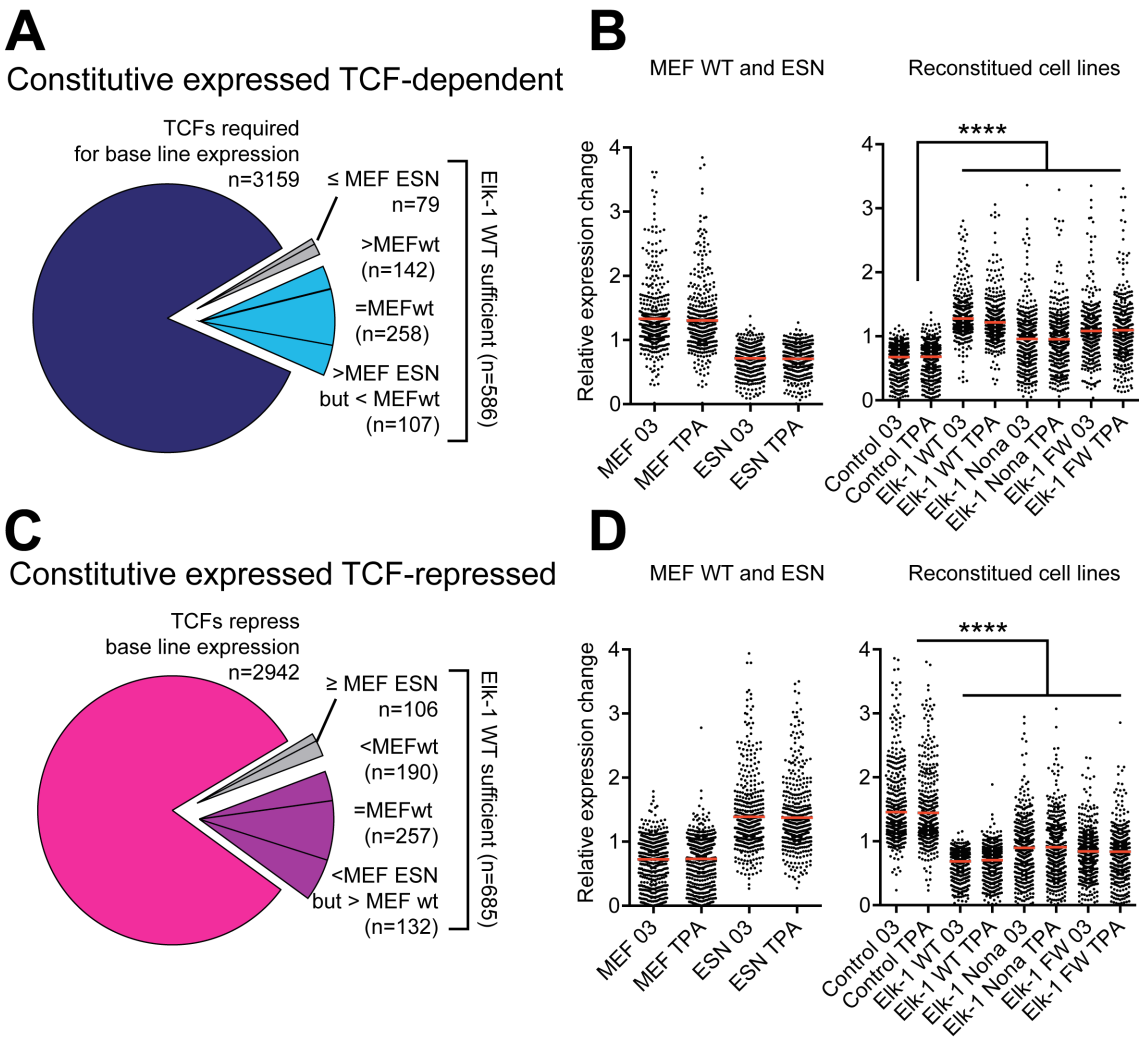


Figure 7.5 Elk-1 is sufficient to control the baseline expression of 20% of the TCF-dependent genes

(A) Analysis of the baseline expression in ESN MEF against wild type MEF for TCF-dependent genes (n=3745). We selected genes whose baseline was impaired in ESN MEF and rescued with Elk-1 (comparison between Elk-1 rescue and control line n=586). From this group we excluded genes showing a baseline expression still impaired if compared to wild type MEF (n=79). **(B)** Relative expression across cell lines for the 507 selected targets. Parental lines uninfected and reconstituted cell lines are displayed. **(C)** Analysis of the baseline expression in ESN MEF against wild type MEF for TCF-repressed genes (n=3627). We selected genes whose baseline expression was repressed in wild type MEF compared to ESN MEF and where the ectopic expression of Elk-1 was sufficient to establish repression in ESN MEF (comparison between Elk-1 rescue and control line n=685). From this group we excluded genes showing a baseline expression still higher if compared to the parental wild type MEF (n=106). **(D)** Relative expression across cell lines for the 579 selected targets. Parental lines uninfected and reconstituted cell lines are displayed. In B and D Friedman test (**** P<0.0001).

ACTIVATED GENES

	TPA-induced genes						Constitutive expression change in ESN (negative)			
	TCF-dep		Elk-1 suff		TCF-ind		Elk-1 suff		Elk-1 not suff	
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
Regulation of transcription	3.5E-14	167	4.2E-03	21	ns	ns	ns	ns	1.2E-06	283
Chromosomal part	7.2E-12	40	3.3E-02	5	ns	ns	8.2E-04	18	ns	ns
Ubl conjugation	6.8E-11	56	9.0E-03.	8	3.5E-04	73	3.9E-06	31	ns	ns
Chromosome, centromeric region	2.1E-10	22	8.8E-03	4	ns	ns	3.8E-02	7	ns	ns
M phase of mitotic cell cycle	3.7E-09	30	1.5E-02	5	ns	ns	3.6E-05	16	ns	ns
Condensed chromosome	1.7E-07	18	6.3E-02	3	ns	ns	6.6E-04	10	ns	ns
Chromosome organization	6.6E-07	41	ns	ns	ns	ns	1.4E-04	23	3.3E-03	58
Negative regulation of biosynthetic process	1.3E-06	42	1.9E-02	7	ns	ns	6.6E-02	16	ns	ns
Microtubule cytoskeleton	8.4E-06	36	9.3E-02	5	2.5E-03	59	5.1E-04	23	ns	ns
Negative regulation of transcription	2.5E-05	35	3.6E-02	6	ns	ns	ns	ns	ns	ns
STAT signaling pathway	6.9E-05	18	ns	ns	5.6E-02	22	ns	ns	ns	ns
Negative regulation of MAP kinase activity	1.3E-03	6	ns	ns	ns	ns	ns	ns	ns	ns
response to organic substance	6.0E-03	35	1.2E-02	8	1.2E-02	64	ns	ns	ns	ns
MAPK signaling pathway	6.5E-03	20	ns	ns	1.7E-06	51	3.5E-02	12	ns	ns
Cyclin, N-terminal	1.8E-02	5	5.9E-03	3	8.7E-03	8	ns	ns	ns	ns
Mitotic sister chromatid segregation	3.9E-02	4	ns	ns	ns	ns	7.3E-03	4	ns	ns
sexual reproduction	4.3E-02	25	ns	ns	7.7E-02	46	7.0E-03	18	ns	ns
Schwann cell differentiation	5.4E-02	3	8.4E-04	3	ns	ns	ns	ns	ns	ns
Chromosome condensation	6.6E-02	4	ns	ns	ns	ns	1.3E-02	4	ns	ns
DNA packaging	6.7E-02	9	ns	ns	ns	ns	1.8E-05	12	ns	ns
M phase of meiotic cell cycle	8.2E-02	8	ns	ns	ns	ns	8.5E-04	9	ns	ns
Cell maturation	ns	ns	ns	ns	ns	ns	7.0E-03	7	ns	ns
Regulation of cellular response to stress	ns	ns	ns	ns	ns	ns	7.0E-03	7	ns	ns

REPRESSED GENES

	TPA-repressed genes						Expression change in ESN (positive)			
	TCF-dep		Elk-1 suff		TCF-ind		Elk-1 suff		Elk-1 not suff	
	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n
small GTPase mediated signal transduction	3.8E-06	21	ns	ns	6.7E-01	17	ns	ns	1.3E-01	41
morphogenesis of an epithelium	3.9E-06	17	ns	ns	ns	ns	1.9E-01	8	1.4E-02	34
embryonic organ development	6.8E-04	16	ns	ns	7.5E-01	15	6.1E-02	12	8.1E-02	40
regulation of epithelial cell proliferation	7.2E-04	8	ns	ns	ns	ns	ns	ns	ns	ns
adherens junction	1.2E-03	10	ns	ns	5.9E-01	7	ns	ns	6.0E-05	31
skeletal system development	3.6E-03	16	9.8E-04	4	1.5E-01	25	7.2E-04	19	ns	ns
regulation of cell shape	5.0E-03	6	ns	ns	ns	ns	ns	ns	ns	ns
small GTPase regulator activity	7.2E-03	13	ns	ns	2.2E-01	19	1.3E-01	10	2.5E-02	41
Hypertrophic cardiomyopathy (HCM)	7.3E-03	8	2.1E-03	3	6.8E-01	5	3.2E-01	4	3.5E-02	19
Valine, leucine and isoleucine degradation	7.6E-03	6	ns	ns	ns	ns	ns	ns	1.9E-02	13
actin filament	9.6E-03	11	ns	ns	ns	ns	4.9E-02	10	9.5E-08	50
regulation of ERK1 and ERK2 cascade	2.2E-02	3	ns	ns	ns	ns	ns	ns	ns	ns
positive regulation of developmental process	3.3E-02	11	ns	ns	9.9E-01	8	2.2E-01	9	ns	ns
regulation of cell	4.3E-02	3	ns	ns	ns	ns	ns	ns	ns	ns
cardiac cell differentiation	4.3E-02	4	2.4E-02	2	ns	ns	ns	ns	1.3E-01	8
neural tube development	4.8E-02	6	ns	ns	ns	ns	ns	ns	2.7E-01	14
cell projection organization	3.5E-01	10	2.2E-02	3	9.5E-01	16	2.6E-02	16	1.7E-03	61
cell adhesion	4.7E-01	15	ns	ns	ns	ns	3.0E-02	24	3.7E-12	132
regulation of chondrocyte differentiation	ns	ns	ns	ns	ns	ns	7.3E-04	5	ns	ns

Figure 7.6 GO analysis of the TCF-dependent Elk-1-dependent targets

Each group of genes was analysed using DAVID. The top half of the table shows genes whose expression is activated either following TPA or where their constitutive expression relies on the TCFs. TCF-dep (red, first column on the left) are genes whose induced expression following TPA requires the TCF (n=871). The Elk-1 suff (orange, second column from the left) are genes induced by TPA, TCF-dependent were Elk-1 is sufficient to re-establish inducibility in ESN MEF (n=92). TCF-ind (yellow, third column from the left) are genes induced by TPA in both wild type MEF and ESN MEF (n=2027). Elk-1 suff (pink, second column from the right) are genes unaffected by TPA whose expression requires the TCFs and Elk-1 is sufficient to re-establish expression (n=507). Elk-1 not suff (blue, first column from the right) are genes unaffected by TPA which baseline expression is TCF dependent but Elk-1 is not sufficient (n=3159). The bottom half is organised in the same way as the active genes but considering functional terms enriched in genes whose expression is repressed either following TPA or their constitutive expression or repression is affected in the ESN MEF compared to wild type MEF: TCF-dep n=531; Elk-1 suff n=16; TCF-ind n=1559; Elk-1 suff n=579; Elk-1 not suff n= 2942. Functional classes enriched with a minimum p-value of 10^{-3} are highlighted in red.

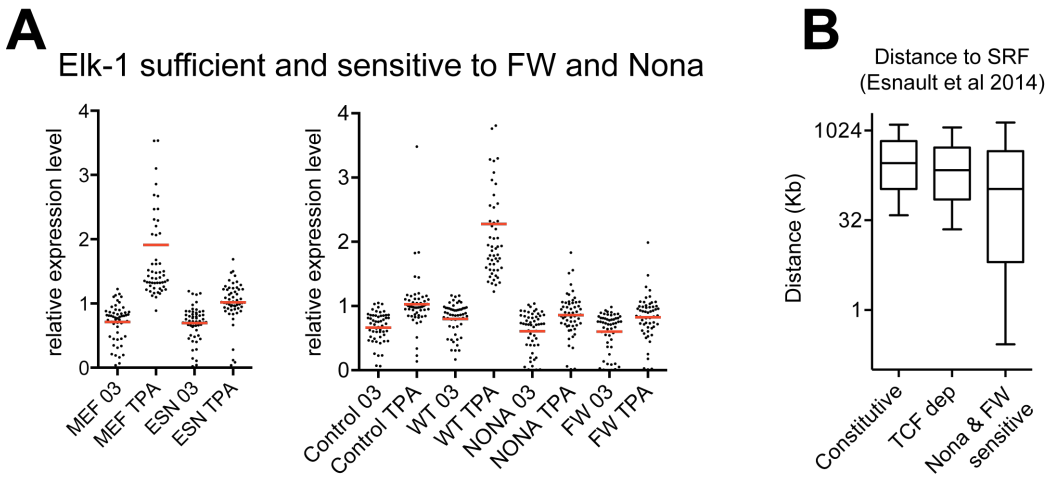


Figure 7.7 Selected group of Elk-1 dependent TPA induced and sensitive to FW and Nona mutations

(A) Relative expression across cell lines of genes TCF-dependent, Elk-1 sufficient and statistically impaired in Elk-1 FW and Nona. (B) Analysis of the distance of SRF to constitutive, TCF-dependent and the selected group of genes using the SRF loci coordinates defined in Esnault et al. 2014.

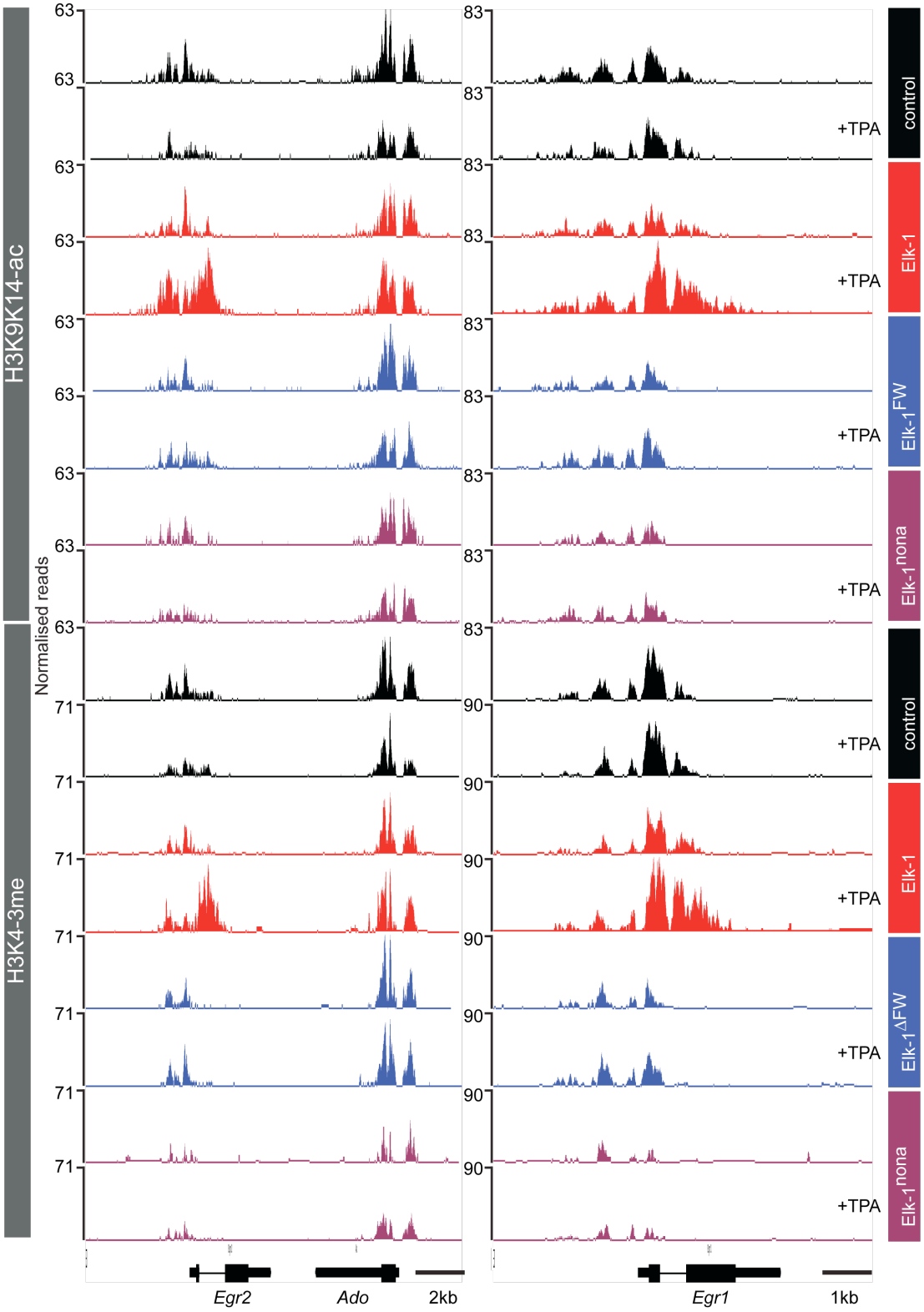


Figure 7.8 H3K4me3 and H3K9K14ac ChIP-seq tracks on Egr2 and *Egr1*.

Top half of the figure display the ChIP-seq signal across *Egr1* and *Egr2* only for the reconstituted ESN MEF for H3K9K14ac ChIP-seq signal while the bottom half shows H3K4me3 ChIP-seq signal. The data displayed shows all reads normalised to the depth of sequencing therefore not normalised to total histone occupancy.

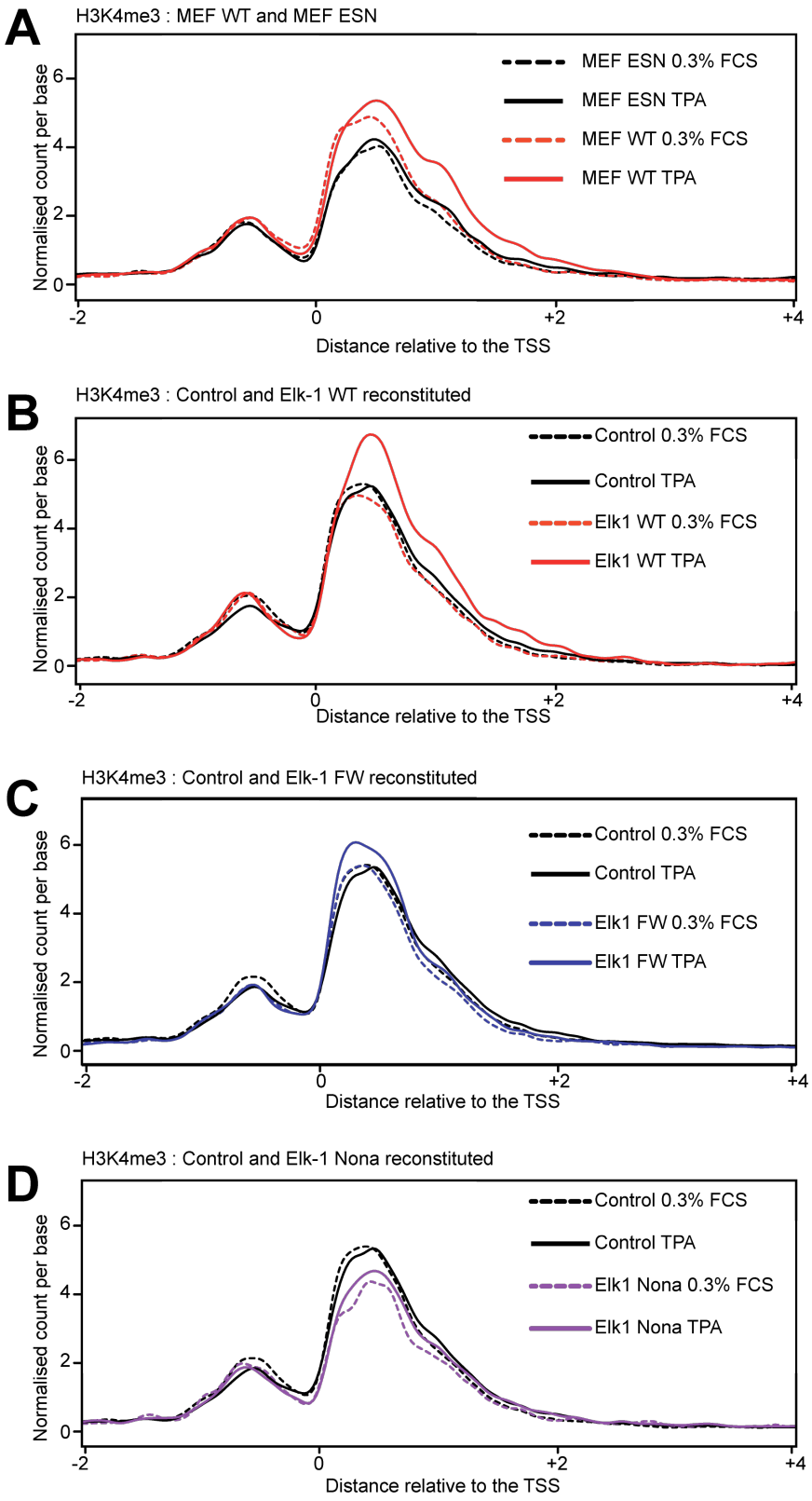


Figure 7.9 H3K4me3 metaprofiles

(A to D) Density plot of the selected targets centred on the TSS of each gene. The data displayed shows all reads normalised to the depth of sequencing therefore not normalised to total histone occupancy.

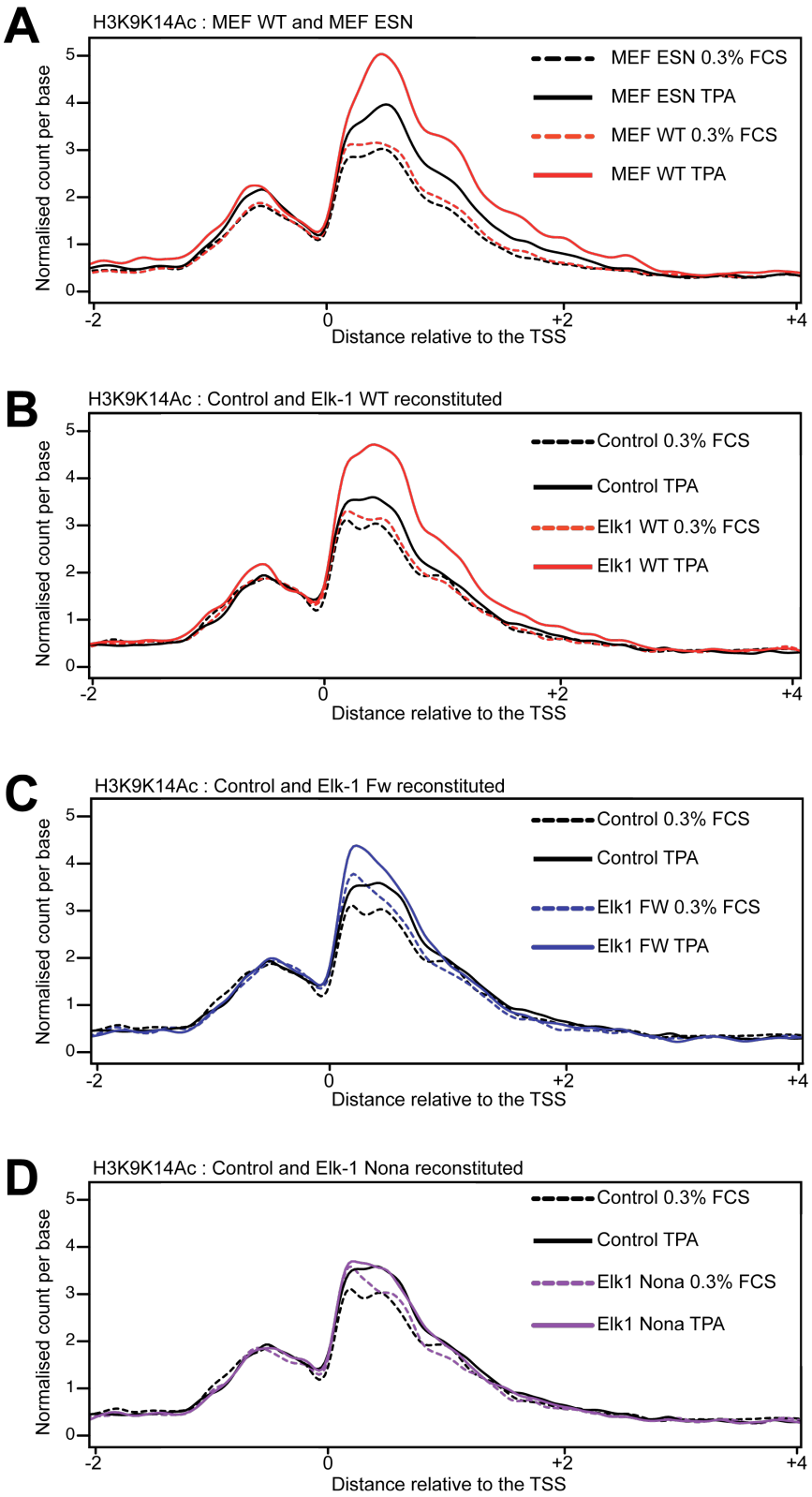


Figure 7.10 H3K9K14ac metaprofiles

(A to D) Density plot of the selected targets centred on the TSS of each gene. The data displayed shows all reads normalised to the depth of sequencing therefore not normalised to total histone occupancy.

Chapter 8. Discussion

8.1 Outline

This thesis explores mechanisms controlling signal-dependent transcription from different angles. Many mechanisms affecting TF-dependent gene activation are considered in the context of the SRF transcription network.

The first three chapters of the results focused on how the SRF co-factor MRTF is controlled once in the nucleus and how this in turn controls transcriptional activation.

o Using genomic approaches it was shown that nuclear MRTF activities are influenced by nuclear actin. Unbiased selection of genes confirmed that actin-dependent gene repression is an MRTF specific phenomenon. Direct interference with the MRTF-actin complex in the nucleus is sufficient to activate SRF-MRTF target genes.

o Nuclear actin directly inhibits MRTF-DNA binding but not Myocardin. The RPEL motifs are the primary regions involved in this mechanism of regulation. Differences within the RPEL1-2 unit between MRTF and Myocardin account for their diverse regulation. Actin-dependent MRTF-DNA binding inhibition occurs dynamically after Rho activation. This is probably due to changes in filament assembly.

o The mechanism of MRTF-dependent gene activation was investigated by assessing steps encompassing Pol II recruitment, escape and productive elongation. MRTF-DNA binding is sufficient for Pol II recruitment and escape. Nevertheless, its disassociation from actin is required for productive transcription to occur. Thus, nuclear actin seems to control mRNA synthesis at steps beyond Pol II recruitment and escape.

The second part of the thesis exploits the Ras/MAPK signalling pathway and the activation of the TCF co-regulators. The immediate-early response in combination with cells lacking all three TCFs provides a valuable system to study how extracellular signals, which target transcriptional activation domains, initiate changes in the chromatin environment.

- Using genomic approaches to assess TCF-dependent gene expression it was possible to describe the transcriptional signature left by the TCFs in response to TPA.
- TCFs control one third of the transcriptional response elicited by Ras-Erk activation. TCFs control genes involved in transcriptional regulation and cell cycle progression. Cells lacking all TCFs show a marked defect in proliferation and a cell cycle block in the G2/M transition. One of the TCFs, Elk-1, although unable to re-establish a correct cell proliferation, is sufficient to control one tenth of the TCF-dependent genes.
- TCF-dependent gene activation in response to TPA is achieved through the establishment of a defined chromatin signature. The signal induced chromatin marks described completely relies on the TCFs. Chromatin changes in response to cues can occur independently from transcriptional activation and rely uniquely on the TCFs' DNA binding and activation domain. The recruitment of the transcription machinery induces a spread of both H3K4me3 and H3K9K14Ac. In addition, the analysis of the H3K27me3 mark demonstrated that Elk-1 activation domain is required in resting conditions to maintain a permissive chromatin landscape.
- MLL3, CHD2, AuroraB and SMYD2 are chromatin modifiers required for *Egr1* and Fos productive transcription. Down-regulation of each of these greatly impairs chromatin marks at *Egr1* promoter without affecting Pol II recruitment.

8.2 Mechanisms governing MRTF nuclear functions

8.2.1 Actin in control of MRTF activity

Nuclear actin impairment of transcription is an MRTF-specific phenomenon

Actin is a negative regulator of MRTF activity. Previous studies showed that actin could also control MRTF-dependent transcriptional activation, but did not address the generality and specificity of this phenomenon (Vartiainen et al. 2007). Confinement of MRTF into the nucleus using Leptomycin B (LMB) does not cause target genes activation. Direct interference with MRTF-actin interaction is required in order to induce productive transcription at MRTF-target genes. The diverse compounds used to dissect MRTF functions have also been shown to affect other

factors. Besides inducing MRTF nuclear accumulation, the Crm1-specific inhibitor LMB allows nuclear retention of additional factors including, among others, NFκ-b, HDACs and Smad proteins. In addition LMB is known to also affect RNA export (Cuevas et al. 2005). Equally, depolymerizing agents such as Latrunculin B (LatB) and Cytochalasin D (CD), besides affecting MRTF activity, have been reported to affect other transcription networks such as the mechanoresponsive YAP-TAZ pathway and NFκ-b (Dupont et al. 2011; Zhong et al. 2013; Kustermans et al. 2005). For these reasons we considered it necessary to run an unbiased search for genes induced by CD regardless of LMB. Genes fulfilling these criteria are spatially associated to SRF-MRTF binding and functionally relate to genes involved in cytoskeletal dynamics. In addition, LMB showed a negligible effect across the genome. Genes constitutively active and linked to SRF showed no changes in expression when cells were treated with LMB. Thus, MRTF nuclear accumulation using LMB is not sufficient to activate productive transcription of SRF-MRTF target genes. The clear implication of these observations is that nuclear actin prevents MRTF from engaging in productive transcription.

As described in Chapter 2, despite the clear impairment of SRF-MRTF target expression following LMB treatment, it is possible to observe a mild activation especially of highly induced genes. Given that this LMB-dependent induction is sensitive to LatB treatment, leads back to MRTF-dependent nuclear activities. This suggests that the transcriptional inhibited state is not absolute. The dynamic behaviour of the MRTF-actin interaction could cause this partial activation. Changes in the relative amounts of MRTF and actin, besides affecting nuclear accumulation, might shift this condition from inactive to active.

Actin may also play a role in regulating nuclear MRTF activity following growth factor stimulation. The MRTF-actin interaction recovers rapidly, within minutes of serum stimulation (Miralles et al. 2003; Vartiainen et al. 2007), and MRTF-DNA binding also decreases even when MRTF remains nuclear (see below).

Cell-type specific regulation of MRTF nuclear activities

Actin might also control MRTF nuclear activities in those cell lines where MRTF is constitutively nuclear. Cell lines such as the Cancer Associated Fibroblast (CAF), MDA-231 (a human breast cancer cell line) and eye fibroblasts, despite

presenting constitutively nuclear MRTF, are still responsive to CD treatments (Medjkane et al. 2009 and unpublished data by Charles Foster and Cynthia Yu-Wai-Man). Thus, it is likely that the actin-dependent mechanisms affecting MRTF nuclear functions are employed by specific cell types in order to control MRTF activity.

Cell lines such as the CAFs are characterised by pronounced stress fibers and more focal adhesion, factors reminiscent of changes in actin filament assembly (Calvo et al. 2013). Relative amounts of monomeric and filamentous actin are characteristic of each cell type and behaviour and could directly affect MRTF activities and subcellular localisation. It is therefore intriguing to consider the MRTF localisation and activity as a potential readout for cellular behaviours.

MRTF activation and mechanosensing

Within the 441 CD induced genes described, gene targets include components of the actin cytoskeleton and factors involved in contraction and cell motility. This signature extensively overlaps with the SRF-MRTF specific set defined by our lab in a recent publication (Esnault et al. 2014). The SRF-MRTF signalling pathway seems to be a crucial component in mechanosensing and transduction in response to Rho-signaling activation (Esnault et al. 2014). In addition to SRF-MRTF, also YAP and TAZ were reported to translocate into the nucleus and activate genes in response to Rho stimulation (Dupont et al. 2011; Yu et al. 2012). Indeed several SRF-MRTF targets are thought to also be controlled by YAP/TAZ including Ctgf, Cyr61 and Ankrd1 (Calvo et al. 2013). Nevertheless CD inhibits YAP activities (Zhong et al. 2013). Therefore the transcriptional signature observed in response to CD stimulation is not due to YAP/TAZ activation.

8.2.2 Actin controls MRTF-DNA binding

Actin directly and specifically controls MRTF-DNA binding activity

Under resting conditions, nuclear accumulation of MRTF is sufficient for its DNA binding (Vartiainen et al. 2007). However, the data presented in this thesis show that high levels of G-actin are sufficient to cause disassociation of MRTF from SRF targets. Indeed actin directly impairs MRTF-DNA binding specifically though the RPEL motifs (see Chapter 3). MRTF mutants, which are unable to interact with

actin, besides being constitutively nuclear, efficiently bind target loci and are unaffected by increases in actin concentration. Similarly Myocardin, an MRTF relative with low actin binding affinity, does not show DNA binding regulation following increases in actin concentration. Analysis of MRTF-Myocardin chimeras indicates that RPEL1 and 2 are critical for the MRTF DNA binding regulation, while RPEL3 is not sufficient. Therefore, due to the differences in actin binding, Myocardin and MRTF DNA binding are controlled in different ways. In addition, DNA pull-down assays performed using MRTF recombinant proteins show a striking reduction of binding activity when monomeric actin was titrated into the assay. This suggests that no further proteins are required for actin to inhibit MRTF-DNA binding. In addition, it was not possible to observe any recovery of monomeric actin on the MRTF-SRF-DNA complex. This indicates that MRTF-actin binding is incompatible with MRTF-SRF-DNA complex formation. In order to draw a full picture of this mechanism of regulation, MRTF mutants lacking the entire N-terminal domain or with point mutations at the critical residues for actin binding should give insights into the sequences involved in this inhibition.

Are MRTF-actin and MRTF-DNA interaction mutually exclusive?

If MRTF-actin binding is incompatible with MRTF-SRF-DNA complex formation, how can MRTF nuclear localisation be sufficient for DNA binding?

MRTF nuclear accumulation following LMB treatments allows the association with target loci across the genome. Yet, MRTF target genes are not activated unless the MRTF-actin complex is severed. The binding intensity under this condition is 50-60% reduced if compared to the transcriptionally competent condition CD. Since more than 90% of the cells treated with LMB show nuclear MRTF, the reduced ChIP signal cannot be ascribed to heterogeneity of the sample. Furthermore reduced ChIP signal was also observed in cells selected to express MRTF linked to the NLS signal. This implies that the ability to interact with basal level of actin affects the degree of MRTF-DNA binding.

Given these observations it is possible to suggest two scenarios: either MRTF-actin binding is incompatible with MRTF-SRF-DNA complex formation or the MRTF-actin complex associates less efficiently with target loci. So far we have failed to see any actin binding at SRF-MRTF target loci performing ChIP

experiments. In addition, as discussed in the previous section, we were unable to observe any recovery of monomeric actin on the MRTF-SRF-DNA complex using purified components. Therefore MRTF-actin binding seems incompatible with MRTF-SRF-DNA complex formation. Thus the reduced ChIP signal reflects the reduced availability of free MRTF.

Recently it has been shown that ChIP experiments could be modelled using known chemical reaction rate theory (Poorey et al. 2013). It is therefore conceivable to see the MRTF ChIP-seq signals as an indication of the likelihood that MRTF binds the given target (Figure 8.1A). Considering that increases in actin concentration (e.g. using LatB) completely abolish MRTF-DNA binding, the binding intensity observed in LMB conditions might directly reflect an imbalance between MRTF and actin relative concentrations. Changes in nuclear actin filament assembly could therefore shift the MRTF-actin complex equilibrium, favouring or disfavouring MRTF-DNA binding. I am currently establishing a TIRF-based assay in order to assess how actin influences MRTF-DNA binding on-off rate.

MRTF-DNA binding as negative feedback mechanism in Rho-activation

MRTF-DNA binding changes after Rho activation. Following an initial peak, 15 to 20 minutes after serum stimulation, MRTF-DNA binding drops by 50%, in parallel with the recovery of the MRTF-Actin FRET as previously reported (Vartiainen et al. 2007). Rho-induced actin filament assembly is followed by disassembly and by an increase in monomeric actin concentration (Vartiainen et al. 2007). It is therefore captivating to consider actin as its own negative regulator given that MRTF directly controls its transcription. In addition, considering that MRTF-DNA binding is the primary cause of Pol II recruitment and escape, impairment of MRTF-DNA binding would be the most efficient way to dampen the transcriptional response (see following section for discussion). The development of assays that are able to directly measure the exact concentration of nuclear monomeric actin following different cues or in different cellular contexts will provide additional information.

A model for MRTF-DNA binding inhibition

An essential question rising from these observations is how does this work? How does actin, through the N-terminal RPEL motifs, directly affect MRTF-DNA binding association? In the MRTF N-terminal half, about 100 amino acids downstream of the last actin-binding unit (RPEL3) is the B1 region, an element required for ternary complex assembly (Zaromytidou et al. 2006). Although we have no indications supporting a possible interaction between actin and this region, it is intriguing to consider a potential role for monomeric actin in occluding the B1 motif. As shown in Figure 8.1A, different mechanisms could be proposed. Firstly, actin could directly interact with the B1 region independently of the upstream RPEL motifs. This mechanism is unlikely given that RPEL1 and 2 are required for the actin-dependent DNA binding inhibition. Further models might involve cooperative binding between the RPELs-actin complex and the B1 region – either directly, or through additional actin recruitment (Figure 8.1A).

Recent structural studies conducted by the lab have established the spatial architecture of the actin-RPEL domain complex (Mouilleron et al. 2011). The structure revealed a pentavalent complex where five actins were associated with the three RPELs. Structural analysis including longer MRTF N-terminal sequences might reveal further actin binding with sequences downstream of the RPEL3. In addition, *in vitro* competition assays between actin and peptides harbouring the B1 sequence would be highly informative in combination with DNA pull-down assays and TIRF-based experiments.

8.2.3 Nuclear MRTF and transcription activation

MRTF-DNA binding is not sufficient for productive transcription

MRTF-DNA binding is sufficient for Pol II recruitment and escape (Figure 8.1B). Comparison of CD and LMB stimulated cells shows that the amount of Pol II recruitment is directly proportional to the MRTF ChIP-seq signal. In addition, MRTF-DNA binding, as a consequence of either CD or LMB stimulation, allows recruitment of pausing factors NELF and DSIF.

Despite the apparently normal Pol II profiles, under resting conditions MRTF-DNA binding is not sufficient for productive transcription. My results indicate

that the inhibited state of Pol II involves reduced CTD Ser2 phosphorylation. Disassociation between MRTF and actin is essential in order to establish the correct pattern of modification displayed at Pol II CTD (Figure 8.1B). Therefore the inhibited state of Pol II somehow precludes productive transcription and mature RNA accumulation.

RNA synthesis is the engine that pushes Pol II forward, and nuclear Run-On and R-loop ChIP experiments show that, even following LMB stimulation, travelling Pol II synthesises RNA. Therefore co-transcriptional RNA degradation must occur. I have attempted to identify and target components of the exosome with the aim of assessing whether RNA can accumulate following LMB stimulation. However the experiments showed high variability and I could not reproducibly identify any potential target. A further investigation, optimising the screen and including additional enzymes known to either stabilise or associate with newly synthesised RNA, is going to be essential to further elucidate the basis for the observed defect.

It is widely accepted that in eukaryotes the steps of RNA maturation including capping, splicing and 3' end cleavage and polyadenylation occur co-transcriptionally ahead of mRNA export (M. J. Moore & Proudfoot 2009). The recruitment of RNA processing factors to the Pol II CTD is what is thought to control pre-mRNA maturation. Several RNA processing factors have been shown to specifically interact with phosphorylated Pol II CTD (see Chapter 1 for a full introduction about CTD modifications and their role and Table 1.3 for CTD interacting proteins).

Under the unproductive LMB condition the escaped Pol II is associated with Ser5 phosphorylation but lacks all the other modifications analysed such as Ser2-P, Ser7-P, Tyr1-P and Thr4-P. The appearance of Ser5-P is consistent with efficient PIC formation and escape of Pol II. Ser5 is known to be a target of CDK7 kinase, which is part of the TFIIH complex (Shiekhattar et al. 1995; Glover-Cutter et al. 2009; M. Kim et al. 2009; Akhtar et al. 2009). TFIIH is essential to induce the open complex transition that precedes Pol II promoter escape (Pan & Greenblatt 1994; Fishburn & Hahn 2012). In addition, MRTF nuclear accumulation allows a correct redistribution of SPT5 and NELF eviction suggesting that, even though no RNA is being detected, p-TEFb might be correctly recruited. Recently it has been shown that p-TEFb preferentially phosphorylates Ser5 and the pausing factors NELF and DSIF (Baumli et al. 2008; Czudnochowski et al. 2012). Previously it was thought

that subunit CDK9 of p-TEFb was the main Serine 2 kinase in metazoan. Its role in Ser2 phosphorylation was proposed because flavopirridol, a potent CDKs inhibitor, was clearly impairing Pol II elongation and p-TEFb activity (Rahl et al. 2010). However several pieces of evidence have shown that CDK9 mainly phosphorylates Pol II CTD at Serine 5 and its activity is not required for elongating Pol II *in vitro* (for an extensive review see Eick & Geyer 2013). More recently CDK12 was proposed to be the main Ser2 kinase in metazoan (Bartkowiak et al. 2010; Bösken et al. 2014). CDK12 preferentially phosphorylates Serine 2 when the Pol II CTD is readily phosphorylated on Serine 7 (Bösken et al. 2014). My results are consistent with this notion as the loss of Serine 2 phosphorylation is accompanied by diminished Serine 7 phosphorylation. I am currently investigating the distribution of CDK12 across *bona fide* SRF/MRTF targets using ChIP experiments and different CDK12-antibodies.

A similar result was recently described by Brookes and colleagues (Brookes et al. 2012). The authors described a set of genes controlled by polycomb and associated with a Pol II uniquely phosphorylated on Serine 5. Consistent with my study, they did not see any phosphorylation at Serine 7 and Serine 2. Similarly to my observations they did not see any accumulation of mature mRNA coming from this group of polycomb-controlled genes. In addition, a direct relationship between polycomb and Pol II occupancy was observed leading the authors to suggest a direct co-dependency between the status of Pol II and the association of polycomb to target genes.

In addition Hargreaves and colluegues, while studying the expression of Toll-like receptor (TLR)-inducible genes in macrophages, identified several genes preassociated with Ser5 but not Ser2-phosphorylated Pol II (Hargreaves et al. 2009). Pol II associated to these genes generate full-length unspliced transcripts that are quickly degraded. Following activation phosphorylation of Pol II at Ser2 allows accumulation of mature transcripts.

We are currently investigating the distribution of Ser7, Tyr1 and Thr4 phosphorylation across the defined SRF-MRTF targets using genomic approaches. Following LMB stimulation these modifications were found impaired across the *Acta2* gene. The full description of the pattern of modifications displayed at Pol II CTD will be crucial in understanding the potential mechanism affecting Pol II productive transcription.

How does this work? As discussed previously, nuclear actin mainly affects MRTF-DNA binding and so far we have failed to see any direct interaction between MRTF and actin at target genes. This observation implies that actin must indirectly, through MRTF, affect the correct modification of Pol II CTD. Given the direct relationship between MRTF-DNA binding and Pol II recruitment, it is tempting to speculate that a less persistence DNA association might therefore be insufficient for the recruitment of defined transcriptional kinases (Figure 8.1B). Again the development of *in vitro* assays to measure the transcriptional outcome given an unstable or persistent MRTF-DNA binding will provide further insights.

This model also needs to take into account how actin directly affects modification of the MRTF polypeptide. Notably, MRTF nuclear accumulation is accompanied by extensive phosphorylation (Miralles et al. 2003). Much evidences collected in the lab have shown that this modification is directly under the control of actin and can affect transcription. Indeed LMB treatment does not induce MRTF phosphorylation (unpublished data collected by F. Miralles, R. Pawlowski and Richard Panayiotou). In addition it has been shown in the lab that Flavopiridol, a potent CDK inhibitor, affects MRTF phosphorylation suggesting a direct involvement of transcriptional CDKs in this process (Richard Panayiotou unpublished data). It is therefore possible to propose that following LMB stimulation actin is able to maintain MRTF in an un-phosphorylated state by continuous recycling through target promoters. Little is known about the contribution of the MRTF phospho-residues in transcriptional activation. It will be interesting to see whether phosphorylated MRTF works as a binding platform for factors involved in gene activation.

MRTF activation does not alter Pol II travelling index

Besides Pol II initiation, additional steps including open complex formation and escape were shown to be under the control of TFs (Rahl et al. 2010; Kouzine et al. 2013). The rate of initiation and escape is supposed to determine the activity of a given gene. TFs can therefore control initiation, escape or both simultaneously. It has been proposed that escape is the rate-limiting step at the majority of the human genome (Core & Lis 2008). Considering this model, one might expect to see a major difference in escape-rate if comparing constitutively expressed and

highly inducible genes following acute stimulation. Constitutive and MRTF-inducible genes showed similar Pol II distribution on average, and the activity of MRTF did not affect the rate of escape when compared to resting conditions. Indeed the probability of escape did not change upon induction with either CD or LMB, implying that initiation and escape are directly coupled. Accumulation of Pol II at the 5' region of any gene reflects a rate limiting escape into the body of the gene and is none other than an indication of active transcription. Corroborating this hypothesis is the observation that in resting conditions genes with a high pausing index are also showing a higher baseline of expression. It is not possible to exclude the possibility that in defined contexts the rate of escape could be specifically affected by defined TFs while others are responsible for maintaining Pol II associated to target genes. We recently reported that in response to serum stimulation, active nuclear MRTF is required for both Pol II recruitment and activation at half of its targets (Esnault et al. 2014). The remaining set of genes was shown to only rely on MRTF for a post-recruitment step. Activation of signalling pathways using serum shock is therefore required to activate additional TFs with the ability to recruit Pol II in the absence of MRTF. Analysis of the interplay between multiple TFs in establishing Pol II elongation rate deserves further investigation.

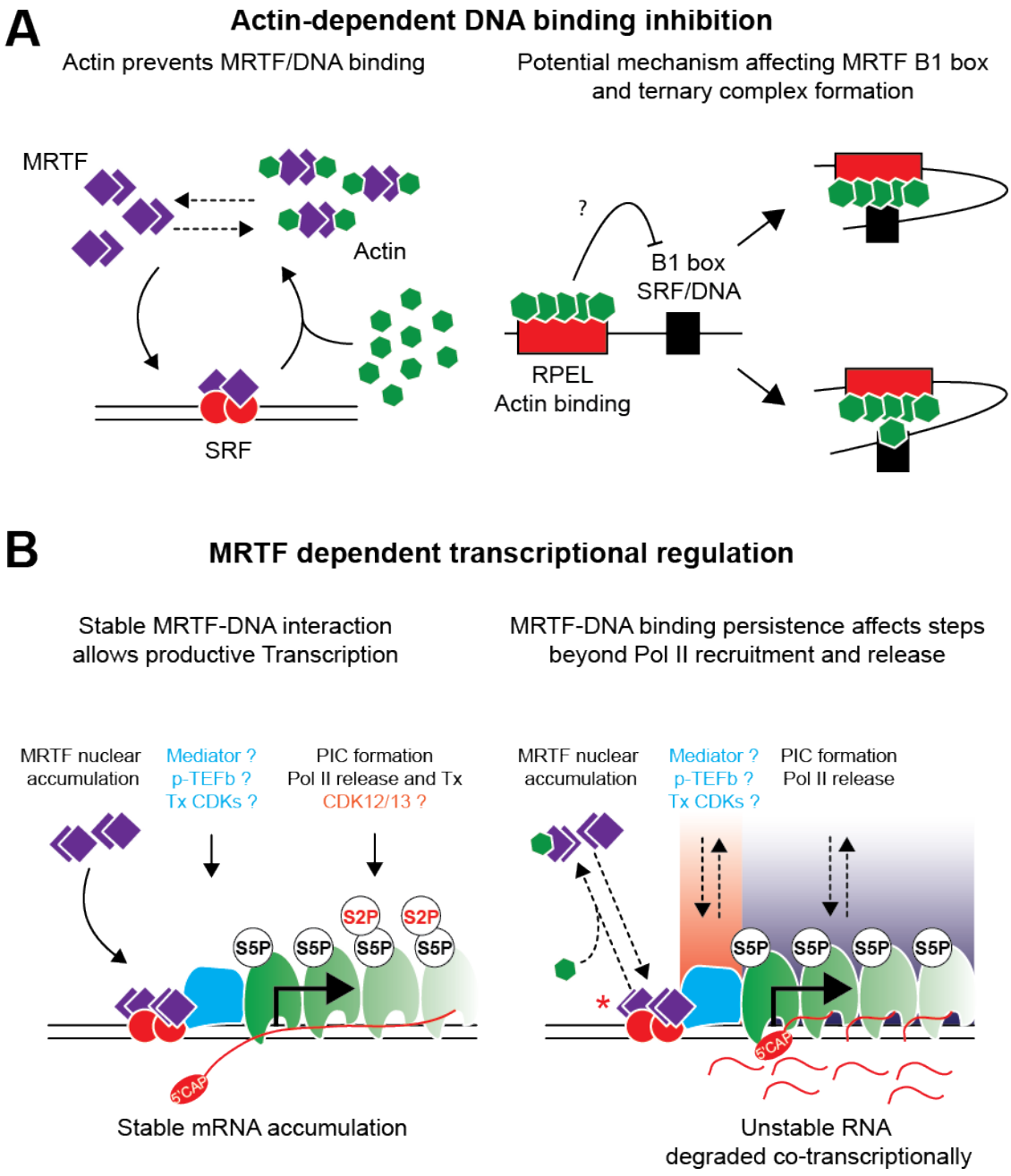

A
Actin-dependent DNA binding inhibition
Actin prevents MRTF/DNA binding
Potential mechanism affecting MRTF B1 box and ternary complex formation
MRTF
Actin
SRF
?
B1 box
SRF/DNA
RPEL
Actin binding
B
MRTF dependent transcriptional regulation
Stable MRTF-DNA interaction allows productive Transcription
MRTF-DNA binding persistence affects steps beyond Pol II recruitment and release
MRTF nuclear accumulation
Mediator ?
p-TEFb ?
Tx CDKs ?
PIC formation
Pol II release and Tx
CDK12/13 ?
S2P
S5P
5'CAP
Stable mRNA accumulation
MRTF nuclear accumulation
Mediator ?
p-TEFb ?
Tx CDKs ?
PIC formation
Pol II release
*
Unstable RNA degraded co-transcriptionally

Figure 8.1 Actin controls MRTF nuclear activities

(A) Actin (green) prevents MRTF (purple) DNA binding and association with SRF (red) at target genes. The impairment of MRTF-DNA binding is concentration dependent and partial MRTF-DNA association might reflect an imbalance between MRTF and actin concentrations. As this phenomenon of inhibition requires intact RPEL motifs a potential mechanism might involve the occlusion of the B1 motif by monomeric actin assembling different repressive complexes. **(B)** MRTF DNA binding is sufficient for Pol II recruitment and escape. In absence of actin interaction, the stable MRTF-DNA binding allows a correct phosphorylation of Pol II CTD and concomitant accumulation of transcript. On the other hand unstable MRTF-DNA interaction, controlled by G-actin, allows Pol II recruitment and escape, but the correct phosphorylation of Pol II CTD is affected. Possibly, under these conditions, MRTF is unable to recruit defined transcriptional kinases. The produced RNA must therefore be co-transcriptionally degraded as Pol II moves along the gene body.
(*) MRTF-Actin interaction also affects modification of the MRTF polypeptide. In particular the accumulation of MRTF into the nucleus following LMB is not accompanied by its phosphorylation. Unphosphorylated MRTF might be unable to recruit specific transcriptional kinases required for Pol II productive transcription.

8.3 Establishment of chromatin signature in response to Ras-Erk activation

8.3.1 The TCF-dependent transcriptional signature

Ras-dependent gene activation is governed by the TCFs

Ras activation is a major cause of proliferation, invasion and metastasis. More than 20% of human cancers are accompanied by Ras activating mutations (for a review see Pylayeva-Gupta et al. 2011). Oncogenic active Ras is known to stimulate a plethora of targets promoting cell cycle entry and proliferation, possibly using different TFs. In particular, AP-1 activity has been widely associated with fibroblast transformation in culture and in the progression of skin tumours driven by Ras (Granger-Schnarr et al. 1992; Lloyd et al. 1991; Saez et al. 1995). Activation of Ras-Erk is a critical event in melanoma, colorectal and renal carcinoma but the degree to which Ras-dependent transformation is dependent on the SRF network is widely unexplored. In particular, there has been no systematic characterization of the TCF-dependent transcriptional response. The analysis of TCF null fibroblasts allowed us to characterise the TCF-dependent gene repertoire, and to assess the role of TCF signalling in the Ras-Erk dependent transcriptional response.

Activation of Ras-Erk results in differential expression of about 5000 genes in wild type MEFs, among which about 3000 are upregulated and 2000 downregulated. The depletion of all TCFs affects about 1000 out of the 3000-upregulated genes and 500 out of the 2000-downregulated genes. In addition the depletion of the TCFs results in differential expression of constitutive expressed genes. Among these, 3745 show reduced expression and 3627 show enhanced expression in TCF null fibroblasts when compared to wild type cells. Genes that are positively controlled by the TCFs are mainly involved in transcription regulation and in the regulation of different aspects of the cell cycle. By contrast genes repressed in a TCF-dependent manner are significantly involved in cell differentiation, tissues and organ development.

Not all TCF-dependent genes need to be direct targets, as they could also be indirect. As I am going to discuss in the next section, the block in cell cycle progression may also influence gene expression in TCF null fibroblasts.

TCFs are required for cell cycle progression and proliferation

Embryonic fibroblasts lacking TCFs show a marked defect in proliferation and cell cycle progression, with a slower growth rate and a block in G2/M of the cell cycle. This observation is consistent with the de-regulation of genes involved in cell cycle regulation. Recently it has been shown that Erk2 and Elk-1 are able to integrate extracellular signals to support ESCs proliferation and prevent their differentiation (Göke et al. 2013). It is therefore tempting to suggest that the TCFs are required for efficient cell cycle progression.

It is important to consider potential indirect effects on gene expression caused by a block in cell cycle. It is widely accepted that several immediate-early genes (IEGs) can regulate cellular differentiation, proliferation and lineage specificity (Fowler et al. 2011). IEGs include several transcription factors, such as c-fos, Egr-1 and AP-1 family members, which could subsequently regulate secondary gene cascades enabling the expression of genes involved in proliferation and cell cycle progression.

Ectopic expression of Elk-1 alone, despite being sufficient to control several TCF-dependent IEGs (Costello et al. 2010), is insufficient to re-establish correct proliferation and cell cycle progression in TCF null cells (data not shown, Cyril Esnault personal communication). Consequently, gene expression changes upon exogenous Elk-1 expression are not likely to reflect cell cycle changes. As I will discuss hereafter, Elk-1 is sufficient to control only a small percentage of the TCF-regulated genes. This might indicate why we failed to successfully rescue the proliferation defect of TCF null fibroblasts.

Despite the functional redundancy that has been reported between Elk-1 and SAP-1 in thymocyte positive selection (Costello et al. 2010), it is conceivable that TCF activities are gene specific. Sap1, Net or a combination of all three TCFs might be required in order to maintain the expression of genes involved in proliferation. In addition Elk-1 and SAP-1 are mostly redundant for those functions, such as thymocyte development, where SRF is also required. Only a defined

subset of SRF-dependent genes are regulated in response to MAPK activation and TCF functions (Esnault et al. 2014). TCFs can act redundantly with other ETS proteins independently of SRF (Hollenhorst et al. 2011). It has been shown that SRF is required for IEGs expression yet is dispensable for proliferation of Embryonic stem cells (Schratt et al. 2001). It is therefore conceivable that the TCFs act redundantly only for the control of SRF-dependent targets while having non-redundant specific functions on other genes.

Elk-1 dependent gene expression

In order to investigate the mechanisms involved in the TCF-dependent gene regulation we used retroviral transgenesis to re-express Elk-1 variants in TCF null fibroblasts. Elk-1-mediated gene activation is dependent on multiple phosphoacceptor sites and two hydrophobic residues, both of which are required to recruit Mediator. TCF null fibroblasts were reconstituted with retroviruses expressing Elk-1 wild type, Elk-1 Nona (a variant in which all phosphoacceptor sites have been changed to alanine) or Elk-1 FW (mutant in which the two hydrophobic residues have been deleted).

The ectopic expression of the Elk-1 wild type protein showed that Elk-1 is sufficient to control about 10% of the TCF-dependent Erk-activated genes (92 out of 871 genes) and about 15% of the TCF-dependent constitutively expressed genes (507 out of 3745 genes). Elk-1 FW and Elk-1 Nona were both unable to active target genes induced by Erk activation. On the other hand, Elk-1 FW and Nona could substitute Elk-1 wild type for the expression of the constitutive genes. This observation implies that Elk-1 DNA binding, rather than its ability to be phosphorylated and to recruit Mediator, is required for the expression of these genes. Potentially Elk-1 DNA binding could favour the recruitment of additional activators at these genes cooperating either directly or indirectly with them. We are currently planning to analyse TF DNA binding motifs at these specific targets in order to see whether there is a specific co-occurrence with Elk-1 binding.

Only a few TPA-repressed genes appeared to be Elk-1 dependent. Greater changes were recorded comparing constitutive expressed genes in wild type and TCF null fibroblasts. Elk-1 is sufficient to suppress about 600 genes whose expression is enhanced in TCF null compared to wild type fibroblasts. Elk-1 was

reported to suppress genes by recruiting co-repressor complexes once de-phosphorylated or specifically sumoylated (S.-H. Yang et al. 2003). Our results show that Elk-1-repressed genes are mostly unaffected by Erk activation. This observation suggests that Elk-1 DNA binding is sufficient to suppress the expression of defined genes. Similarly to Elk-1 dependent constitutive genes, ectopic expression of both Elk-1 FW and Nona showed to be sufficient to establish gene suppression. The mechanism of gene suppression does not seem to be influenced by Erk activity. A potential mechanism might involve the direct competition with other TFs for the same DNA binding site.

The TCFs are part of the large family of ETS transcription factors (see chapter 1.6). ETS members show similar and partially overlapping DNA binding specificity (Wei et al. 2010). It would be important to compare the DNA binding distribution of Elk-1 with the binding of other ETS member in order to gain further insight into this mechanism of gene suppression.

8.3.2 TCFs as anchoring points for signal-induced chromatin changes

TCFs are required for signal-induced chromatin changes

Activation of target TFs, modification of chromatin, and direct recruitment of components of the transcriptional machinery are all potentially regulated by extracellular signals (see Chapter 1). Using the SRF-TCF network, I investigated the relationship between signalling, Elk-1 phosphorylation, chromatin changes and transcriptional machinery recruitment occurring at the *Egr1* gene.

Activation of transcription was shown to correlate with defined chromatin changes including: H3S10 phosphorylation, H3K4 trimethylation, H3K9K14 and H3K27 acetylation and H4K16 acetylation (see chapter 1.4). The rapid and transient activation of the Ras-Erk pathway elicits chromatin modifications associated with transcription activation. A localised phosphorylation of histone H3S10 at the SRF/TCF binding site is the first modification to occur in response to TPA, followed by enhanced acetylation of lysine residues 27, 9 and 14 of the histone H3 and lysine 16 of histone H4. The appearance of these chromatin marks coincides with the recruitment of the Mediator complex, followed by Pol II recruitment at the TSS. H3K4me3 was the most stable mark, being detectable at the SRF-TCF site in resting conditions, but spreading towards the TSS when Pol II

was recruited, as did H3K9K14 acetylation. All the modifications analysed did not occur in TCF null fibroblasts, and were restored by the expression of Elk-1. Thus, the TCFs are required to both mediate the establishment of chromatin changes and recruit the transcriptional machinery at *Egr1* in response to Ras-Erk activation.

Elk-1 DNA binding and phosphorylation induce defined chromatin changes

Previous studies have shown that transcription activation by activated Elk-1 involves the specific recruitment of the Mediator complex (Balamotis et al. 2009). The TCFs, in particular Elk-1, are equipped with a well-characterised activation domain harbouring several phosphorylation sites and two key hydrophobic residues (see chapter 1.6), both of which are required for recruitment of the Mediator complex. We studied the relationship between Elk-1 phosphorylation and chromatin modification by using Elk-1 mutants either lacking all phospho sites (Nona) or the hydrophobic residues (FW). These mutants allowed us to uncouple Elk-1 phosphorylation from active transcription.

Ectopic expression of Elk-1 FW, despite being unable to recruit the transcriptional machinery and activate transcription, was sufficient to enhance H3S10 phosphorylation, H3K9K14 acetylation and H4K16 acetylation at the SRF/TCF binding site following Erk activation. On the other hand TCF null fibroblasts reconstituted with the Elk-1 Nona mutant could not induce H3S10, H3K9K14 and H4K16 modifications. Elk-1 FW could only restore, in response to signal, modifications in the vicinity of the SRF/TCF binding site, while downstream modifications at the *Egr1* TSS were left unchanged. Therefore, chromatin modification can occur in absence of transcription activation at the TCF/SRF site. In summary, it is possible to distinguish two types of signal-induced chromatin modifications: one directly enhanced by Elk-1 phosphorylation alone and others reflecting the recruitment of the transcription machinery.

In contrast, both Elk-1 FW and Nona were sufficient in resting and stimulated conditions to enhance H3K4me3 at the SRF/TCF binding site at *Egr1* promoter. As mentioned earlier, H3K4me3 is the least changing modification in response to signal. Following stimulation, H3K4me3 shows the greatest changes at the TSS when Pol II is recruited. Taken together, these observations suggest that Elk-1 DNA binding is sufficient to enhance H3K4me3 proximal to its DNA binding

site while the transcriptional machinery is required to induce changes at the downstream TSS.

Elk-1 DNA binding and phosphorylation are not sufficient for H3K27 acetylation in response to Erk activation. This modification occurs exclusively at the SRF/TCF binding site and coincides with Mediator recruitment. These observations imply that Mediator recruitment is essential to enhance H3K27 acetylation. To date, Mediator has been reported to exhibit HAT activity only in yeast (Lorch et al. 2000). However, several chromatin-modifying complexes such as SAGA have been reported to act in synergy with the Mediator complex in both yeast and mammals (X.-F. Chen et al. 2012). The specific role of H3K27 acetylation besides being enriched at active genes and enhancers is still unclear. The identification of H3K27 specific HATs will provide additional insights.

Mediator and Pol II provide further chromatin modifications

The chromatin signature described at the SRF-TCF binding site is directly mediated by TCF-DNA binding and phosphorylation of their activation domain. However the distribution of the chromatin modifications change when transcription occurs, suggesting that Mediator and Pol II provide additional functions able to extend the chromatin modifications. H3K4me3 and H3K9K14Ac are enhanced at *Egr1* TSS following TPA and require the recruitment of both Mediator and Pol II. Both acetyltransferase and methyltransferase have been reported to be either directly or indirectly recruited by the transcribing Pol II (Wittschieben et al. 1999; Wood et al. 2003). The direct relationship between H3K4me3 and active promoters has been observed in several studies and is commonly used as a mark of active promoters (reviewed in Ruthenburg et al. 2007). Recently Roeder and colleagues showed that H3K4me3 stimulates Pol II PIC formation by interacting with the TAF3 subunit of TFIID at active promoters (Lauberth et al. 2013). The spreading of H3K4me3 by Pol II might therefore re-enforce its own activity. This mechanism preferentially occurs at signal or stress induced genes rather than constitutively express genes (Lauberth et al. 2013).

The differential distribution of H3K4me3 at both enhancers and promoter sequences raises the general question of whether this modification is mediated by different enzymes or by position-specific activities of a common methyltransferase.

Given the data presented, MLL3 and SMYD2 are the two potential methyltransferases differentially recruited at the *Egr1* gene. Further experiments including ChIP of the enzymes involved in *Egr1* activation and Co-IP experiments between Elk-1 and the chromatin modifiers will provide further insights.

Functional role of chromatin changes in TCF-dependent gene activation

How do chromatin changes contribute to TCF-dependent gene activation? As described above, it is widely accepted that signal activation induces several chromatin modifications correlating with sites of active transcription (J. Ernst et al. 2011). In particular H3S10 phosphorylation has been implicated in IEGs activation (Mahadevan et al. 1991 and for a review see Sawicka & Seiser 2012). Despite the strong correlation between signal-induced chromatin changes and transcription activation, it is not yet clear which of the steps in the transcriptional cycle these changes are promoting.

It has been reported that the induced de-methylation of the stem cell leukemia (SCL) locus using TALE-LSD1 fusion protein affects nearby gene expression (Mendenhall et al. 2013). Thus chromatin modifications are required for transcription to occur, but in which step are they required for?

To identify chromatin modifiers involved in Erk-mediated gene expression we carried out a medium-throughput siRNA screen (see chapter 6). The siRNA-mediated knockdown of Aurora-B, MLL3, CHD2, Set7, Tip60, Tip49b and SMYD2 inhibits both *Egr1* and Fos expression, and Aurora-B, MLL3, CHD2 and SMYD2 depletion also impairs specific marks at the *Egr1* promoter. In particular Aurora-B knockdown reduces H3S10 phosphorylation while MLL3 and CHD2 reduces both H3K4me3 and H3S10 phosphorylation. SMYD2 knockdown reduced the appearance of H3K4me3 at *Egr1* TSS following TPA stimulation. Therefore different chromatin modifiers contribute to *Egr1* expression. Notably, the reduction in gene expression and chromatin modification is not accompanied by a reduction in Pol II recruitment, suggesting that the chromatin changes might influence steps beyond Pol II recruitment.

It has been proposed that H4K16 acetylation is sufficient to recruit BRD4 and p-TEFb at active promoters (Zippo et al. 2009). Our results show that H4K16 acetylation is not sufficient to recruit p-TEFb. Indeed TCF null fibroblast

reconstituted with Elk-1 FW does not show p-TEFb recruitment in response to Erk activation, whereas H4K16 acetylation is induced. Both chromatin modifications and Mediator recruitment might be required for p-TEFb recruitment and Pol II activation. The analysis of the distribution of components of the p-TEFb complex and BRD4 following the down-regulation of Aurora-B, MLL3, CHD2 or SMYD2 will provide further insights.

A model for TCF-dependent gene activation

We propose that the TCFs work as anchoring points where both chromatin modifier enzymes and the Mediator complex are recruited in response to signals (Figure 8.2). Chromatin changes and Mediator recruitment appears to represent two distinct pathways, both under the control of the TCFs, contributing to transcription activation.

The first pathway requires both phospho acceptor and FW residues of Elk-1 AD in order to recruit the Mediator complex as reported by the Berk lab (Balamotis et al. 2009). In addition our results show that Elk-1 phosphorylation and the FW residues are sufficient for Elk-1-Mediator interaction and Pol II recruitment, even when chromatin modifications are impaired (Figure 8.2 step a). The second pathway is characterised by the establishment of chromatin modifications. H3K4me3 at the SRF/TCF binding site requires Elk-1-DNA binding (Figure 8.2 step b) while Elk-1 phosphorylation allows H3S10 phosphorylation and acetylation of H3 and H4 (Figure 8.2 step c). Chromatin changes at the SRF/TCF binding site are completely independent of the transcription machinery. This might reflect a direct - or indirect - recruitment of chromatin modifiers through the Elk-1 AD. Chromatin modifications at the SRF/TCF binding site are per-se insufficient for transcription activation and might therefore favour a step beyond Pol II initiation (Figure 8.2 step d). Chromatin modifications and Mediator might together favour Pol II release through the recruitment of p-TEFb (Figure 8.2 step d). Ultimately the transcribing Pol II allows H3K4me3 and H3K9K14Ac to spread (Figure 8.2 step e).

As described in Figure 8.2 the chromatin modifiers could be placed in different steps of the activation mechanism. MLL3 and CHD2 could be recruited in resting conditions by the TCFs and methylate H3K4me3 (Figure 8.2 step b). MLL3 is mainly known to induce H3K4me1 but several studies suggest it is also involved

in H3K4me3 deposition (Vicent et al. 2011; Valekunja et al. 2013). Aurora-B might be directly or indirectly recruited by phosphorylated Elk-1 allowing H3S10 phosphorylation at the SRF/TCF binding site (Figure 8.2 step c). H3S10 phosphorylation is followed by acetylation of histone H3 and H4 at different residues. So far we failed to identify potential acetyltransferases and we cannot interpret the interdependency between the various chromatin modifications. Nevertheless all changes require TCF and appeared to be differentially affected following Aurora-B, MLL3 or CHD2 knockdown. Finally SMYD2 might be recruited in a Pol II dependent manner at active promoters leading H3K4me3 to spread into the transcribed sequences (Figure 8.2 step e).

Signalling to TCF as a mechanism to maintain active chromatin

It has been recently reported that Erk2 and Elk-1 genomic distribution in human embryonic stem cells defines active and poised genes (Göke et al. 2013). Genes occupied by both Erk and Elk-1 are associated with cell cycle progression and proliferation, whereas genes associated with Elk-1 alone relate to cell differentiation and commitment and are not transcribed. The H3K27me3 mark, possibly established by the polycomb group of proteins, characterises this latter group. Interestingly, silent Elk-1 genes fall into the same functional categories as the one specifically suppressed by the TCF as described by our RNA-seq data. In this context the mechanism of repression seems to rely exclusively on the DNA binding activity. Indeed the Elk-1 Nona and FW mutants, as the wild type protein, can restore their normal transcription pattern. The direct implication is that this group of genes is suppressed in a signal-independent way. This group of genes opposes to the pro-proliferative targets positively controlled by the TCFs. Therefore, it is possible to speculate that the TCFs are required to maintain a stable cell identity.

Although it is not clear how Elk-1 can be differentially targeted by Erk protein across the genome, a possible implication is that activation of the TCFs provides an antagonistic mechanism to silencing. Intriguingly the Elk-1 Nona mutant, which cannot be phosphorylated, allows accumulation of H3K27me3 across the *Egr1* ORF. This observation was recently confirmed performing ChIP-seq experiments and we are currently analysing the data. A direct implication is that TCF

phosphorylation itself prevents the appearance of H3K27me3. Indeed the Elk-1 FW mutant does not enhance H3K27me3. Therefore the potential antagonistic mechanism, although requiring Elk-1 phosphorylation, does not require active transcription. Recent studies described that Erk activity pulses in growing cells and this mechanism controls the proliferation rate (Aoki et al. 2013). Therefore Erk activity pulses fired spontaneously or through cell propagation could in theory allow TCF activation and chromatin modifications in resting conditions. Indeed H3K27me3 was not seen in resting conditions in any cell line analysed. Therefore the signal-dependent chromatin signature occurring at the SRF-TCF binding site might allow the maintenance of an active chromatin template even in resting conditions. Further studies are required in order to draw a comprehensive picture of this mechanism.

The function of H3K27me3 in this particular context is not established and the involvement of the polycomb group of protein is yet to be determined. It is intriguing to consider recent studies conducted by the Helin lab that described how polycomb is not required to induce silencing whereas it is indispensable to maintain it (Riising et al. 2014). Therefore persistent inhibition might prevent stochastic activations leading to active chromatin modifications. The establishment of repressive states could therefore be indirectly brought about rather than specifically induced. A better undestanding of how H3K27me3 is established at TCF-target genes and the potential role of the polycomb complex in this mechanism is of crucial interest and it will be further investigated.

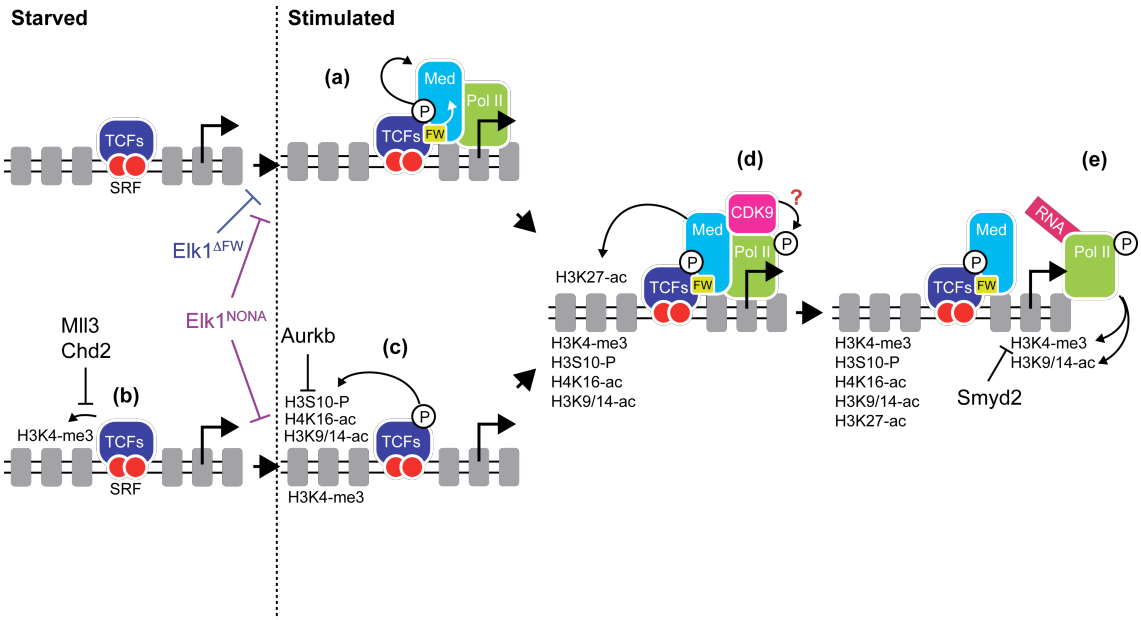


Figure 8.2 Establishment of chromatin signature in response to Ras activation

(a) Activated TCFs recruit the Mediator via direct contact with defined subunits of the complex (Balamotis et al. 2009). **(b)** TCFs through their DNA binding are able to enhance H3K4me3 at *Egr1* promoter. This modification occurs in absence of signal stimulation and transcriptional activation and relies on MLL3 and CHD2 activities. **(c)** Upon stimulation phosphorylation of the TCFs' activation domain enhances H3S10P, H4K16Ac and H3K9K14Ac at the SRF/TCF binding site**. (d)** Chromatin changes and Mediator are both required for p-TEFb recruitment and transcription activation. **(e)** The established signature at the TCF-SRF binding site, consisting in H3K4me3, H3S10P, H4K16Ac, H3K9K14Ac and H3K27Ac (the latter induced only in the presence of the Mediator complex) is then expanded with the spreading of H3K4me3 and H3K9K14Ac towards *Egr1* TSS. This re-shaping of the chromatin signature is entirely dependent on the recruitment of the transcriptional machinery and requires SMYD2 activity.
The maintenance of the chromatin signature in resting conditions seems indispensible to keep a permissive chromatin landscape.

Chapter 9. Materials & Methods

9.1 Chemicals and reagents

The following list shows the most commonly used reagents and materials. Specific reagents used are highlighted in each section.

3MM paper	Whatman
Ampicillin	Sigma
Blasticidin	Invitrogen
Bromophenol Blue	BioRad
BSA	Sigma
Comassie brilliant blue	BioRad
Complete protease inhibitor cocktail tablets	Roche
Cyrochalasin D	Calbiochem
Dimethyl sulfoxide (DMSO)	Fisher Scientific
Dithiothreitol (DTT)	Sigma
Doxyciclin	Sigma
Dynabeads® M-280 Streptavidin	Invitrogen
Dynabeads® Protein G for Immunoprecipitation	Invitrogen
Dynabeads® Rat Anti-Mouse IgM	Invitrogen
Ethidium bromide	Sigma
Fetal Calf serum (FCS)	Invitrogen
Hepes	Invitrogen
Kanamycin	Sigma
Latrunculin B	Calbiochem
Milk powder	Marvel
Optimem	Invitrogen
Orange G	Sigma
Phenylmethyl-sulfonyl fluoride (PMSF)	Sigma
Poly(dIdC)•poly(dIdC)	Sigma
Polybrene	Sigma
Protein marker SeeBlue Plus2 Prestained	Invitrogen
Puromycin	Invitrogen

Spermidine	Invitrogen
TPA	Sigma
Triton X-100	Invitrogen
Trizma-base	Invitrogen
Tween® 20	Invitrogen
Zeocin	Invitrogen

9.2 Expression vectors

Protein expression in mammal cells

MRTF and Myocardin derivatives were cloned into a modified pTRIPZ vector purchased from Dharmacon (TRIPZ Inducible Lentiviral shRNA). Each mutant and derivative was sub-cloned from existing vectors in the lab. Each cDNA was inserted between AgeI and MluI sites using homologous recombination (InFusion, Clontech). Bovine growth hormone polyadenylation (bgh-PolyA) signal, derived from pcDNA™4/TO (Invitrogen) was introduced into the MluI site downstream of the inserted cDNA using homologous recombination (In-Fusion, Clontech).

Elk-1 mutants were introduced into pMYs-IRES-GFP. Each cDNA was collected from existing plasmid generated by Thomas Ross in the Lab and described previously (Cruzalegui et al. 1999; M. A. Price et al. 1995; Marais et al. 1993).

MRTF(fl) and SRF(fl) protein expression in Sf9

MRTF and SRF full-length cDNA were optimised using the Invitrogen GeneArt® services. The original sequence was obtained from the one in use in the lab and already described (Guettler et al. 2008; Miralles et al. 2003; Vartiainen et al. 2007). The optimised sequence included a Histidine tag at the N-terminal part fused to a TAP tag (Tandem affinity purification) and followed by a TEV protease site (Tobacco Etch Virus nuclear inclusion a endopeptidase, aa sequence ENLYFQG). The vector used to express the 6XHis-TAP-TEV-MRTF(fl) and SRF(fl) was a modified pBacPAK (Clontech). From the vector I removed the entire region from NcoI to KpnI. The cDNA was then cloned, following PCR to include restriction sites at the extremity, between KpnI and NotI.

Additional plasmids were used including vectors expressing the viral VSV-G envelop (Clontech) and the SV40 large T (Addgene).

9.3 Bacterial manipulation

9.3.1 Bacterial strains

One Shot Top 10 E. coli chemically competent cells were used for routine cloning. One Shot® Stbl3™ Chemically Competent E. coli derived from the HB101 E. coli strain were used for cloning and expansion of the pTRIPZ and pMYs-IRES-GFP. These vectors harbours long terminal repeats (LTRs) so are highly prone to homologous recombination. The Stbl3 strain is design to minimise recombination of retrovirus and lentivirus vectors harbouring LTR sequences.

9.3.2 Bacterial media

LB media: 1% w/v Bacto-tryptone, 0.5% w/v Bacto-yeast extract, 1% w/v NaCl)
LB Agar: 1% w/v Bacto-tryptone, 0.5% w/v Bacto-yeast extract, 1% w/v NaCl, 1.5% w/v Bacto-agar
SOC media: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4, and 20 mM glucose.

LB liquid media and agar plates were supplemented with antibiotics according to the resistance marker harboured by the plasmid: 100µg/ml ampicillin, 30 µg/ml kanamycin, 25-50 µg/ml Zeocin, 50–100 µg/mL Blasticidin.

9.3.3 Transformation of competent cells

50 µl of chemically competent Top 10 or Stbl3 were thawed on ice and mixed with the DNA for 10 minutes. The mix was then incubated at 42°C for 30 seconds and then quickly transferred on ice for 2 minutes. The bacteria were then incubated with 250 µl of SOC media without antibiotic for 1 hour before plating on agar plate with the desired antibiotic.

9.4 Nucleic acid manipulations

9.4.1 Plasmid DNA purification

5 ml (for small-scale, miniprep) or 200 ml (for large scale, maxiprep) of LB media, including the appropriate antibiotic, were incubated with a single Top10 or Stbl3 colony and grown overnight at 37°C and 200 rpm. Miniprep plasmid DNA was isolated by the CRUK Equipment Park miniprep service while Quiagen Plamid kits were used to purify though anion-exchange column larger quantities of DNA (Maxiprep).

9.4.2 Nucleic acid quantification

The concentration and purity of double stranded DNA and RNA were routinely quantified using a ND-1000 UV/Vis spectrophotometer (NanoDrop Technologies). Samples generated for high-throughput sequencing were quantified and analysed by the Advanced Sequencing facility using the Agilent 2100 Bioanalyzer. This device is a microfluidic-based system able to assess size, quantity and quality of DNA and RNA.

9.4.3 Elecrophoresis

1-2% agarose gels were prepared in 1X TBE (89 mM Tris Base 89 mM boric acid 2 mM EDTA) supplemented with 0.5 µg/ml of Ethidum bromide. Samples were supplemented with 10X loading buffer (0.01 % w/v Orange G, 30 % glycerol in TE pH 8.0). Gels were run using CLP electophoresis tanks at 100-150 V. The reference marker used was the 2-Log DNA Ladder 0.1-10.0 kb (New England Biolabs).

9.4.4 Recombinant DNA techniques

cDNA fragments were generated either by Sequence-specific endonucleolytic cleavage of plasmid or PCR products. In addition complex cloning involving highly recombinant vectors or for long plasmid backbones the homologous recombination-based approach In-Fusion was used (Clontech, technology based on (M. Z. Li & Elledge 2007).

Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify DNA for cloning or to generate biothynilated oligonucleotides for DNA pull down assays using MyTaq Red (Bioline, see later). Specific primers were designed by me and synthesised by Sigma. PCR products were routinely produced using KOD Hot start polymerase kit (Novagen). For Recombinant based cloning techniques according to the manufacturer instruction each primer used was designed to overlap for at least 15 bases with the target sequence (see Clontech web site). For standard molecular cloning techniques primers included restriction sites at the extremities.

Standard PCR reaction:
10ng DNA
1µl of each primer forward and reverse (0.2µM final)
5µl 10x PCR reaction mix
5µl 25 mM $MgSO_4$
0.2mM final each dNTP
1µl KOD Polymerase
Water to a final volume of 50µl

PCR cycling:
Denaturation: 95°C for 5 minutes at first, for each cycle allow 30 seconds
Anneal: 50 to 65°C according to primers and templates, for 30 to 45 seconds
Elongation: 68-72°C according to the type of amplification, for 30 to 60 seconds

Each cycle comprising Denaturation, anneal and elongation was repeated 20 - 40 times according to the quantity needs (NB increases in cycle numbers increase the probability of errors although the KOD polymerase is a proof reading enzyme).

Insertion of DNA fragments into plasmid DNA

For standard cloning procedure digestion of both plasmid and insertion product was conducted according to the manufacturer's instruction (New England Biolabs). Calf intestinal alkaline phosphatase (New England Biolabs) was added

when required to the target vector to avoid self-ligation by removing the 5' phosphate groups. The PCRed and digested DNA fragments were purified with agarose gel electrophoresis and/or MiniElute PCR product purification kit (Qiagen). Ligation was performed using the T4 DNA ligase (NEB) according to the manufacturer's instructions. 2 µl of the ligation mix were used to transform Top10 or Stbl3 strains. For the homologues recombinant cloning approach I followed the manufacturer's instruction using the In-Fuction HD kit (Clontech).

DNA sequencing

DNA sequencing performed by the CRUK Equipment Park sequencing services was used to verify each plasmid generated and used.

9.4.5 List of Plasmid used

pTRIPZ_MRTF-HA$_2$NLS:

Full length MRTF-A ("MRTF-A siRAR") cDNA generated originally by F. Miralles (Miralles et al. 2003; Zaromytidou et al. 2006) and was later modified by R. Pawlowski (Pawłowski et al. 2010). The cDNA was corrected for point mutations introduced originally during PCR and the primary DNA sequence is modified in order to become resistant to siRNA oligos targeting from nt 521 to 532 (sequence AGCTGGTGGAGA changed in AACTAGTAGAAA). HA$_2$NLS was sub-cloned from plasmids encoding MAL-HA originally generated by S. Guettler (Guettler et al. 2008) by digesting MAL-HA with BbsI and XbaI. The HA2 NLS was fused to the siRAR-MRTF-A and then cloned into the pTRIPZ vector using homologous recombination as described above.

pTRIPZ_MRTF-HA$_2$NLSmut:

This plasmid was generated as the pTRIPZ_MRTF-HA2NLS. The MRTF-A siRAR was used as substrate to introduce HA$_2$NLSmut (with point mutations destroying the nuclear localisation signal) originally cloned by S. Guettler (Guettler et al. 2008). The HA$_2$NLSmut was introduced using BbsI and XbaI sites for sub-cloning.

pTRIPZ_MRTF$^{123\text{-}1A}$-HA$_2$NLSmut:

The *pTRIPZ_MRTF-HA$_2$NLSmut* was used as substrate to sub-clone using BamHI sites the RPEL motifs at the N-terminus of the MRTF$^{123\text{-}1A}$ originally cloned by S. Guettler obtaining an MRTF$^{123\text{-}1A}$ siRAR.

pTRIPZ_MC-HA$_2$:

Mouse Myocardin variant A obtained originally by F. Miralles by RT-PCR from mouse cardiac RNA and subcloned by S. Guettler in pEF as fusion to 2 C-terminal HA tags. The cDNA was clones by homologous recombination into pTRIPZ vector. The cDNA was described in (Guettler et al. 2008).

pTRIPZ_N12MC-MRTF-HA$_2$:

cDNA described in (Guettler et al. 2008). AA 2-148 of MRTF-A was exchanged for AA 1-94 of MC. The cDNA was cloned by homologous recombination into pTRIPZ vector.

pTRIPZ_N123MC-MRTF-HA$_2$:

cDNA described in (Guettler et al. 2008). AA 2-187 of MRTF-A was exchanged for AA 1-133 of MC. The cDNA was cloned by homologous recombination into pTRIPZ vector.

pMY_Elk-1 WT

The cDNA was originally obtained by Thomas Ross in the lab and described previously (Cruzalegui et al. 1999; M. A. Price et al. 1995; Marais et al. 1993). The original human Elk-1 cDNA was fused at its N-terminus with a Flag tag leaving the ATG starting codon of Elk-1. The ATG codon was modified in GTG by PCR during the cloning of Elk-1 in the retrovirus expressing vector pMY-IRES GFP between XhoI and NotI.

pMY_Elk-1 FW

The cDNA was originally obtained by Thomas Ross in the lab and described previously (Cruzalegui et al. 1999; M. A. Price et al. 1995; Marais et al. 1993). The original human Elk-1 FW cDNA was fused at its N-terminus with a Flag tag leaving the ATG starting codon of Elk-1. The ATG codon was modified in GTG by PCR

during the cloning of Elk-1 in the retrovirus expressing vector pMY-IRES GFP between XhoI and NotI. The AA in positions 378 and 379 (corresponding to the FW AA) were deleted.

pMY_Elk-1 NONA

The cDNA was originally obtained by Thomas Ross in the lab and described previously (Cruzalegui et al. 1999; M. A. Price et al. 1995; Marais et al. 1993). The original human Elk-1 cDNA was fused at its N-terminus with a Flag tag leaving the ATG starting codon of Elk-1. The ATG codon was modified in GTG by PCR during the cloning of Elk-1 in the retrovirus expressing vector pMY-IRES GFP between XhoI and NotI. Amino acids in position 324, 336, 353, 363, 368, 383, 389, 417 and 422 were mutated to alanine.

pBACpAK_SRF and pBACpAK_MRTF

The original sequence was obtained from the one in use in the lab and already described (Guettler et al. 2008; Miralles et al. 2003; Vartiainen et al. 2007) and the cDNA was optimised using the Invitrogen GeneArt® services for sf9 expression. The cDNA was cloned through homologous recombination into a modified pBApAK (see previous sections). Each cDNA is fused at its N-terminus to a 6x His tag followed by a TAP tag and a TEV site.

9.5 Mammalian cell culture

NIH3T3 and MEF derived cells at 50% confluence were cultured in DMEM (Gibco, Invitrogen) supplemented with 10% FCS at 37°C and 10% CO2. Cells carrying pTRIPZ vectors were maintained in media supplemented with puromycin at 1µg/ml. To split confluent cells, cells were washed with PBS and then treated with trypsin/versene until they detached (2-3 minutes). Media containing 10% FCS was added and cells were distributed to new dishes. Cells were serum starved in 0.3% FCS DMEM either over night for NIH3T3 or for 48h for MEF cells. For the selection of cell harbouring the desired plasmid, cells were transfected with lipofectamin reagents (see following section) and treated with puromycin (1-2 µg/mL) for 2 to 3 weeks. Cells not transfected were treated in parallel as reference for negative selection of cells not expressing the resistance marker.

9.5.1 Stimulation conditions: Drugs concentration

The conditions used in each experiment included a step of serum starvation (overnight for NIH 3T3 and 48h for MEFs). Cells were then treated for the indicted times using different components:

Leptomycin B (LMB) was used at 50nM. A mix containing warm DMEM with no serum and LMB at 150nM was vortex in glass vials. Following starvation the media was removed from cell plates and a third of the final volume of DMEM and LMB mix was add to the cells. Two thirds of the final volume of warm DMEM without serum was then ad to reach the final concentration of LMB (50nM). This step was performed to ensure homogeneity of cells as was observed that LMB is prone to precipitate in aqueous solutions. The same steps were followed for the 0.3% FCS condition using warm DMEM with no LMB and serum to exclude possible effects caused by the changes in media.

Cytochalasin D (CD) was used at a final concentration of 3μM and the steps of media vortexing and changes were followed as for the LMB condition.

Latrunculin B (LatB) was used at a final concentration of 0.5μM and the steps of media vortexing and changes were followed as for the LMB and CD conditions.

12-O-Tetradecanoylphorbol-13-acetate (TPA) was used at 50ng/ml. For homogeneity warm DMEM with no serum was mixed with TPA and add to the cell layer.

9.5.2 Transient transfection of plasmid and siRNA

NIH 3T3 were transfected using Lipofectamin 2000 reagent (Invitrogen) following the manufacturer's instruction. This procedure was used to generate stable cell lines harbouring pTRIPZ vector. I observed that in order to have efficient expression of the target protein 10cm plates needed to be treated with at least 10-15 μg of plasmid DNA. For plasmid transfection cell were seeded on the day of transfection, 5 to 6 hours were allowed for the attachment of the cells in 10 cm plates and the media was substituted with 4ml of Optimem. 10-15 μg of plasmid DNA was mixed with 0.5 mL of Optimem and 10 μl of Lipofectamine 2000 was mixed with an equal amount of Optimem. After 5 minutes incubation the two mix were combined and incubated at room temperature for 20 minutes. The DNA

lipofectamin mix was then added to the cell layer and left to incubate for 2 hours in the incubator. The transfection mixture was then replaced with warm fresh media.

HEK-GP293 (Clontech) were transfected using Fugene (Promega) under containment conditions. To transfect a 10cm of growing GP-293 557µl of Optimem were mixed to 23µl of Fugene in polypropylene tubes and incubated for 5 minutes at room temperature. In a clean polypropylene 7.5µg of retroviral vector together with 6 µg of VSV-G expressing vector were combined with the Fugene Optipmem mix. After 15-30 minutes of incubation the transfection mix was added drop wise to the GP-293.

siRNA oligos were transfected into MEF cells using a reverse protocol to maximise performance using the RNAimax (Invitrogen). The siRNA oligos were purchased from Dharmacon as ON-TARGET plus pool of 4 or as single ssRNA oligos. The dry oligo pellet was dissolved in 1x siRNA buffer from Dharmacon to a final concentration of 20µM. The assays were performed in different plates or dishes according to the type of experiment.

For the siRNA screen 24 well plates were used. 62.5µl of Optimem were mixed with 0.625µl of siRNA oligo (stock solution 20µM) and 62.5µl of Optimem were mixed with 0.75µl of RNAiMax (Invitrogen). The mix were incubated for 5 minutes at room temperature and then mixed and left to incubate at room temperature for 20-30 minutes. During this incubation cells were detached and diluted in media to reach a concentration of 20000 cell/ml. The 125µl of siRNA and RNaiMax were mixed directly in the well of a 24-well plate with 500 µl of cell suspension. The transfection was carried for 72 hours. Cells were then starved for two days and treated for 30 minutes with TPA (50ng/ml).

To perform siRNA ChIP assay the quantities were scaled up: 1.5ml of Optimem were mixed with 15 µl of siRNA oligos and 1.5ml of Optimem were mixed with 18 µl of RNAiMax. The two mix were incubated for 5 minutes at room temperature and then mixed for 20-30 minutes. 12ml of cells in suspension at a concentration of 60000 cell/ml were mixed directly in 15cm dishes with the 3ml mix. As above the cell were left for 72h then starved for two days and finally treated with TPA for 30 minutes.

9.5.3 Retrovirus infection and cell sorting

Retroviruses were prepared using HEK-GP293 and Phoenix cells able to produce the virus in big quantity. These cells were cultured as for NIH3T3 and MEF. As low adherent no trypsin/versene was used, instead mild re-suspension with media was enough to detach the cells from the plate. The two stages infection from GP-293 to phoenix and then to target cells was followed to minimise cross infection and hazards. GP-293 are HEK-293 cells that stably express Gag and Pol viral proteins. In order to allow the production of viral particles transfection of the retroviral vector, containing the protein of interest, and the vector harbouring the envelop protein VSV-G is required. Phoenix cells on the other hand are helper-free cells meaning that they express all the constituents in order to produce a viral particle given a retroviral vector equipped with the viral package signal and the transcription/processing elements. Phoenix cells were seeded at $2.5X10^5$ in 6-well plates on the day before the infection (24h after transfecting the Gp-293). The supernatant of GP-293 cells was harvested 48h after transfection and fresh media was added to the cells (see previous section). The supernatant was passed through a 0.45μm filter and supplemented with Polybrene at a final concentration of 5ug per ml. 2ml of filtered supernatant was used to infect each well of the 6-well plate of phoenix. The plates were centrifuged at 1200 rpm 26°C for 90minutes in a centrifuge with swinging buckets. This step was repeated twice with the remaining media. The following day the procedure was repeated with the media harvested from GP-293 (72h after transfection). The infected Phoenix cells were grown and expanded for 2 weeks in containment conditions and freezed.

To infect target cells, such as ESN MEF, Phoenix cells were expanded and grown in 10cm dishes for 48h. The media was harvested and the procedure as described above repeated. ESN MEF were grown in 6-well plate and infected twice in a day with media harvested at 48h and once with media harvested at 72h. Cells were grown and expanded until reaching a 70% confluent 15cm dish.

As ESN MEFs were derived from knockout mice expressing several markers of resistance we used the pMY-IRES-GFP vector to ectopically express Elk-1 derivatives. Stable cell lines expressing the desired amount of protein were collected according to the GFP expression using Fluorescence Activated Cell Sorting (FACS) by the FACS laboratory. Three populations were collected as high,

low and medium expressing according to the GFP level. Cells were expanded and the protein amount compared performing western blots.

9.5.4 Mouse embryonic fibroblast preparation and immortalisation

Robert Nicholas and Victoria Lawson generated ESN MEF from mice Elk-1$^{-/-}$ SAP-1$^{-/-}$Net$^{\partial/\partial}$. After having sacrificed a pregnant mouse at the desired dpc (day post coitum) the uterine horns were dissected and placed into cold PBS. Each embryo was separated from the horns and placental and embryonic sac in petri dishes. The head and red organs were separated from the main trunk. The collected tissue was washed in PBS and placed into a clean 6-well plate. Parts of the head or red organs were used later to collect total genomic DNA for genotypiong. The tissue was then finely minced and placed into glass vials with magnetic stirrer and 2ml of trypsin/versene. Each glass vial was incubated in a water bath at 37°C for 15min under agitation. Following these steps 10ml of warm media supplemented with 10% FCS was added to the shredded tissue and slowly re-suspended up and down several times. Cells were centrifuged at low speed (300xg) for 5 minutes, the supernatant was removed and the cell pellet re-suspended with fresh DMEM 10% FCS media and plated in 10 cm dishes.

MEFs were immortalised using SV40 T antigen. Phoenix cells were grown in 10cm plates and when 70% confluent were transfected using Fugene with 2µg of pBABE-SV40 large T-hygromycin. Supernatant was harvested 48h after transfection, filtered with 0.45µM syringe filter and gave to MEF as medium in 10cm plates with 8µg/ml of polybrene. MEFs were then allowed to reach confluence. After 24h from the infection cells were split and after 6 to 8h hygromycin (Invitrogen) was added at a concentration of 50µg/ml. In parallel some MEF were grown without infection as a control for negative selection.

9.5.5 Cell growth analysis

In the result section we analysed the growth and proliferation of MEF wild type and ESN MEF using three approaches:

Crystal violet staining assay

Cells were seeded at equal concentration in 10cm plates (15-20% confluent) and allowed to grow for 3 days. Cells were fixed in 3.7% Paraformaldehyde (PFA) for 5 minutes and then incubated for 30 minutes with an aqueous solution of Crystal Violet 0.05% (Filter at 0.45µm before use). The plates were washed with double distilled water several times and dried.

Cell Enumeration time course:

Cells were seeded in several plates at 15-20% confluent and harvested with trypsin/versene everyday for 4 days. Cells were fixed with cold 70% ethanol and counted using Fluorescence Activated Cell Sorting (FACS) by the FACS laboratory.

BrdU staining

Cells were treated with 10µM Bromodeoxyuridine (BrdU) for 4h, harvested and then fixed with 1ml of ice cold 70% ethanol under agitation and incubate on ice for 30 minutes. Cells were then centrifuged and washed twice in PBS. The FACS laboratory performed the following steps: Cells were re-suspended in 500µl of 2M hydrochloric acid (HCl) and incubated at room temperature for 20-30 minutes. Cells were pelleted at 2000rpm for 5 minutes and washed twice in PBS and once in PBS-Tween (PBS + 0.1% BSA + 0.2% Tween 20, pH 7.4). Pellet cells were incubated for 20 minutes in the dark at room temperature with 100µl of anti-BrdU antibody (Sigma cat no. B5002, stock 1mM) diluted 1 in 50 in PBS-Tween. Following a wash with PBS-Tween cell pellets were incubated with the secondary antibody anti-BrdU antibody (Becton Dickinson, cat no. 347580) diluted 1 in 200 in PBS-Tween for 20 minutes in the dark at room temperature. Following a wash in PBS-Tween cell pellets were incubated with 50µl of ribonuclease A and 150µl propidium iodide (PI) at room temperature, in the dark for at least 30 minutes.

The distribution of the cell population across the G1, S and G2/M phases were analysed using a flow cytometry system and the data was processed using FlowJo. Several gates were applied through the analysis to select single cells and exclude debrides (see London Research institute technologies web site).

9.5.6 Immunofluorescence

The immunofluorescence approach was performed on coverslip either obtained from 6-well plates, were cells were grown 60-70% confluent, or from

coverslip included in 15cm dishes used to prepare chromatin samples after fixation (see the Chromatin immunoprecipitation section).

Following the desired condition, usually including a step of over night serum starvation and 30 to 1 hour stimulation, cell were fixed with 4% paraformaldehyde for 10 minutes at room temperature (coverslip obtained from 15cm during the preparation of chromatin samples were fixed with 1% final formaldehyde for 10 minutes at 37°C and then incubated with Glycine 250mM final for 5 minutes). Cells were washed 3 times with PBS and then permeabilised with 0.2% Triton-X in PBS for 10 minutes. Non-specific binding was then blocked with blocking solution (10% FCS, 1% fish skin gelatine in PBS) for 1 hour and room temperature. Following this step cells were incubated with the cell side on 50µl drops containing the primary antibody diluted in blocking solution for 1.5 hours in a close humidified environment. Coverslip were briefly washed by dipping in PBS three times and then incubated with the secondary antibody including DAPI and phalloidin. Finally cells were washed three times in PBS, once in double distilled water and then mounted on microscope slides using Mowiol. Cells were incubated in the dark at room temperature to let the Mowiol to dry. The images were taken using a Zeiss Axiovert microscope equipped with a camera. As described in the figure legends throughout the result section the localisation of MRTF was scored as nuclear, cytoplasmic or pancellular in at least 150 cells per slide analysed. The distribution of MRTF was assessed considering where it was found predominantly between the two cellular compartments.

9.6 Protein purification

9.6.1 Expression and purification of recombinant SRF and MRTF

The Protein Purification Facility (PPF) at Lincoln's Inn Fields Laboratories performed the generation of insect cell culture, the isolation of recombinant baculovirus expression vector using the BacPAK system (Clontech) and the infection of target insect cells (Sf9, Spodoptera frugiperda). For an extended protocol see Laboratories 2013.

Sf9 pellet cells, obtained following infection, were kept at -80°C. On the day of protein extraction and purification the pellet was thaw in 10ml of Lysis Buffer

(50mM NaPi buffer pH 7, 500mM KCl, 0.005% Brij-35, 1mM DTT, 5mM EDTA, 1mM EGTA, Protease inhibitions) per gram of cell pellet. The mix was pipetted up and down several times and transferred into a glass beaker, pre-cooled on ice, together with a stir magnetic bar. The lysate was mixed in a cold room for 15 minutes to dissolve clamps. Cells were lysed by passing them twice through a pre-cooled (on ice) “french-press” or dounced 6x20 times on ice. The lysate was then centrifuged at 50000 rpm (256630 rcf) for 40 minutes at 4°C in a Ti70 UZ rotor. During the centrifugation 0.5ml of IgG Sepharose beads, per 2 liters of culture (GE helth care), were equilibrated with the lyse buffer by washing 3 times in 10ml and pelleting them at 1500 rpm for 5 minutes. The quantity of beads can vary according to the amount of protein expressed by the Sf9. The clarified lysate was mixed with the equilibrated beads in batch and let under rotation overnight at 4°C. The day after, beads were pelleted at 1500 rmp for 5 minutes at 4°C and combined in a 50ml falcon (Note: the beads were never let to dry, about 0.5ml of buffer were always left at the bottom of the falcon tube). Cells were washed with 50ml of lyse buffer 5 times. Lastly the pellet of beads was resuspend in 1.5-2ml of lyse buffer and combined with 30µl of AcTEV Protease (Invitrogen) in a 2ml eppendorf tube and let over night at 4°C on a rotating wheel. The day after the supernatant was collected and separated by the beads using a chromatography empty column equipped with a filter at its bottom. The beads were washed with an additional 5ml of lysis buffer, to collect supernatant soaked into the sepharose. The collected protein was concentrated to reach about 0.5ml volume using an ultrafiltration spin column with a cutoff of 30KDa (Vivaspin turbo 15, Sartoriusstedim biotech). The concentrated protein was then subjected to gel filtration chromatography on a calibrated Superose 6 column (GE Healthcare) using the same lysate buffer. 500µl fractions were collected and 10 µl of each were resolved by 4-12% SDS-PAGE followed by Coomassie brilliant blue staining. Fractions showing a good quantity of the desired protein were combined and concentrated again to reach 0.5ml volume. The collected protein was quantified using standard Bradford assay and aliquoted in 0.1 ml Individual UTW Tubes (Thermo scientific) and snap froze in liquid nitrogen. The aliquots were kept in paper boxes either in liquid nitrogen or at -80°C.

9.6.2 Purification of rabbit skeletal muscle actin

This protocol was performed together with Richard Panayiotou and Jessica Diring.

Acetone Powder preparation

Rabbit skeletal muscle actin was purified as previously described (Feuer & Molnar 1948; Spudich & Watt 1971). All the procedure was carried out at 4°C unless stated differently. About 500g of white rabbit leg muscle was minced several times with the help of a mean mincer. The minced muscle was mixed for 10 minutes in 2 litres of 10mM KCl solution, 10 minutes in 2 litres of 50mM $KaHCO_3$ solution and 10 minutes in 2 litres of 1mM EDTA solution. The homogenate was separated in half (each part corresponding to 250gr of starting material) and each part was then washed twice for 5 minutes in 3.5 litres of water and then brought to room temperature. The homogenate was briefly washed in 3.5 litres of cold acetone and then washed three times for 10 minutes each in two volumes of acetone at room temperature. These last washes were done using a magnetic stirrer to keep the homogenate mixing during the 10 minutes incubation. At this stage the homogenate was dried on 3MM Whatman paper in a fume hood and the acetone powder was then stored at -80°C.

Actin extraction

10gr of acetone powder were hydratated in 200ml of ice cold G-Buffer (5mM Tris-HCl pH 8, 0.2mM $CaCl_2$, 0.2mM ATP, 0.5mM DTT, 0.15mM PMSF) for 1 hour on ice. The homogenate was then filtered through 4 layers of plain muslin and the filtrated collected as the fraction containing actin. This step was repeated once more with 100ml G-Buffer and the two filtrated pooled together. The Filtrated were cleared from insoluble materials by centrifugation at 27000 xg at 0°C in a Beckman JLA- 16.250 rotor in a Beckman Coulter Avanti J-25 centrifuge.

The filtrated was adjusted to 1 mM ATP, 2 mM MgCl2 and 50 mM KCl and incubated under slow stirring for 2 hours in order to polymerise the monomeric actin in solution. The actin binding proteins, contaminating the sample, were removed by adding solid KCl to 600mM and mixed under vigorous stirring for 1.5 hours. The polymerized actin was collected by centrifugation at 34,000 rpm for 1.5h

in a Beckman Coulter Optima L-100 XP ultracentrifuge with a Beckman Type 50.2 Ti rotor. The pellet made of F-actin was resuspended in G-Buffer containing 0.5mM ATP and 1mM DTT to reach about 30ml. The mix was Dounce homogenized on ice between 60 and 100 times. The mix was then dialysed against 2 litres of G-Buffer for 3 days.The concentration of ATP during dialysis was reduced on the last day to 0.2mM. Lastly monomeric actin was separated from filamentous actin and insoluble material by centrifugation at 100000 xg for 30 minutes. Finally actin was stored in aliquots after snap-freeze at -80°C.

Preparation of Latrunculin B-actin

The final actin stock was kept in solution with Ca^{2+}. In order to convert the Ca^{2+}-actin into Mg^{2+}-actin was dialysed overnight into Mg^{2+}-G-buffer (2 mM Tris-HCl pH 8.0, 0.3 mM MgCl2, 0.2 mM EGTA, 0.2 mM ATP, 0.5 mM DTT). A 5-fold molar excess of latrunculin B (LatB; Calbiochem), quickly mixed to the dialysed actin and incubated overnight at 4°C. To remove uncomplexed actin, a 20x initiation buffer (2 M NaCl, 60 mM MgCl2, 10 mM ATP) was added and the mixture was incubated for 2 hours at 4°C. Following this step filaments were precipitated by centrifuging the sample at 2000000 xg for 15 minutes at 4°C in a Beckman TLA 120.2 rotor in a in a Beckman TL-100 tabletop ultracentrifuge. The supernatant was aliquoted and stored at -80°C.

9.7 Western-blotting

Proteins samples were collected by harvesting cells cultured in 6-well plates directly by adding 200µl of 2x laemmli buffer (100 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 5% β-mercaptoethanol, bromophenol blue). Samples were collected by scraping the wells with CytoOne Cell Scraper (220 mm, 11 mm Fixed Blade from Starlab) and sonicated for 1 minute in 1.5ml eppendorf tubes in a BioRuptor water bath sonicator. Proteins were separated according to their size using sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). NuPAGE Novex Bis-Tris 4-12% gradient gels (Invitrogen) were used following the instruction provided by the manufacturer. 20µl of protein lysate were loaded per lane and the electrophoresis was carried out at 150-200 volts in MOPS SDS running buffer (Invitrogen). The

SeeBlue Plus2 Prestained Standard (Invitrogen) was used as reference in each experiment.

The primary antibodies used were: SRF (G20 sc-335, Santa Cruz), HA (High affinity peroxidase conjugated 3F10, Roche), phospho-p44/42 ERK1/2 (Cell Signalling), Flag (F7425, Sigma), MRTF-A (C-19 sc-21558, Santa Cruz), Actb (C4 sc-47778, Santa Cruz). The secondary antibodies were all purchased from Dako and use 1:10000.

9.8 DNA pull down assay

Preparation of magnetic DNA affinity beads

Two Biotynilated DNA fragments, 120 bp long, were generated by PCR using as templates the wild-type c-fos promoter plasmid, pF711, containing 711bp of the 5' flanking sequence (Treisman 1985) and the pFosΔSRF plasmid (Hill & Treisman 1995a), where the SRE was mutated to create an MCM1 binding site were SRF cannot bind anymore.

The PCR reaction was carried out using MyTaq Red (Bioline) preparing 50 µl of reaction mix as follows:

50ng Template plasmid
1x NH_4 Buffer
5 µl $MgCl_2$
1 µl dNTP mix
1 µl of Forward Biotynilated primer (Sequence: CGCACTGCACCCTCGGTG)
1 µl of Reward primer (Sequence: GGTCCCCCCCCCAGAACAACAGGGA)
1 µl of RedTaq
and H_2O to reach 50µl

PCR cycling:
Denaturation stage: 95°C for 5 minutes at least as first stage, for each cycle allow 45 seconds
Anneal stage: 65°C, for 45 seconds
Elongation stage: 72°C for 45 seconds

Each cycle comprising Denaturation, anneal and elongation was repeated from 30 times.

The PCR reaction was scaled up to 96 independent reactions to reach a total of 4.8ml. The product was phenol-chloroform extracted by mixing the sample 1:1 with a saturated phenol:chloroform:isoamyl alcohol solution (25:24:1, Sigma). The sample was then centrifuge at maximum speed for 5 minutes in a tabletop centrifuge. The upper aqueous phase was moved to a clean tube and the described steps repeated twice again (phenol:chloroform:isoamyl alcohol mix, spin and phase separation). Traces of phenol were removed by mixing 1:1 the samples (aqueous phase) with chloroform and centrifuged for 5 minutes. The upper aqueous phase was then transferred to a clean tube and subject to ethanol precipitation. The sample was mixed with 3M NaOAc ph 5.2 (1/10 of the sample volume), and 2.5-3 volumes of 100% ice cold Ethanol. The mix was incubated on dry ice for 1 hour and then centrifuged at maximum speed for 30 minutes in an Eppendorf 5415 R tabletop centrifuge at 4°C. The pellet of DNA was washed twice with 500µl of 70% ethanol solution in water. Finally the pellet was re-suspended in 500µl of TE solution (TRis-HCl 10mM, EDTA 1mM).

The biotin-labelled dsDNA was coupled to streptavidin magnetic beads (Dynabeads® M-280 Streptavidin, Invitrogen). Beads were washed twice in DW Buffer (20 mM Tris-HCl pH 8.0, 2M NaCl, 0.5mM EGTA, 0.03%NP40). The biotynilated dsDNA was coupled to the washed beads in 0.5x DW Buffer for 30 minutes at room temperature on a turning wheel (each mg of dynabeads were coupled to 250 pmol of dsDNA, considering the 120bp dsDNA used each mg of dynabeads were coupled to 20µg of DNA). The dsDNA coupled to beads were washed five times in DW buffer and finally once in TE-NP40 (10mM Tris-HCl, 1mM EDTA, 0.02% NP40). Beads were equilibrated in 1x Buffer G (10mM Tris-HCl pH 8, 20mM HEPES pH 8, 1.5mM $MgCl_2$, 100mM NaCl, 0.2mM EGTA, 10mM Potassium glutamated, 0.02% NP40, 10mM DTT) and then used for the assay.

DNA pull down assay

The DNA pull down was carried out in G Buffer in the presence of 3mM Sperimidine and 20ng/µl of polydI-dC (Sigma). The equilibrated Dynabeads

coupled to the dsDNA were mixed with recombinant SRF and incubated under agitation for 20 minutes at 30°C. The beads were then washed 5 times with 500µl of G Buffer and used as substrate for MRTF. MRTF binding and washes were carried out as for SRF. The amount of both SRF and MRTF were estimated by performing a titration of both in order to obtain a good signal keeping the negative control MCM1 devoid of unspecific interaction. Finally pure actin was mixed in G Buffer at increasing concentrations with the washed beads carrying SRF and MRTF and incubated for 15 minutes at 30°C. The beads were washed 5 times in G Buffer and the bound proteins collected in SDS sample buffer for western Blotting.

9.9 Gene expression analysis by quantitative Q-PCR

Cells were grown in 6-well plates and staved either over night (NIH 3T3) or for 48 hours (MEF wild type and ESN MEF) and then stimulated as described in each experiment. RNA was extracted using GenElute Mammalian Total RNA Miniprep Kit following the manufacturer instruction (Sigma). The extracted RNA was eluted from columns in 50-70µl of ultra clean double filtered water. 5µl of total RNA were treated with 1µl of DNAseI (Ambion, Invitrogen) and 1µl of DNaseI buffer (provided with the enzyme). The mix was incubated in PCR strips or PCR plates. The PCR strips and plates were incubated in termo-cycler for 15 minutes at 37°C. The reaction was stopped by adding 1µl of 25mM EDTA solution and incubation at 65°C for 10 minutes. The DNaseI treated mix of 8µl (5µl of initial RNA, 1µl of DNaseI, 1µl of DNase I buffer and 1µl of EDTA solution) was then used to synthesise cDNA using SuperScript III (SSIII) First Strand Synthesis system and random hexamer primers (Invitrogen) as follow. 1µl of Ransom hexamer and 1µl of dNTP mix were add to the 8µl sample and incubated at 65°C for 5 minutes and quickly chilled on ice. 10µl of cDNA synthesis mix (2µl of 10x Buffer solution, 4µl of 25mM $MgCl_2$, 2µl of 100mM DTT, 1µl of RNaseOUT from Ambion, 0.5µl of SSIII and 0.5µl of water) were add to the treated samples and mixed well by pipetting up and down. The cDNA synthesis was performed as follows: 10 minutes at 25°C, 50 minutes at 50°C and 85°C for 5 minutes. The samples were diluted with ultra clean water 1:10.

SYBR-Green based real-time PCR (Invitrogen) was assembled by mixing 8µl of reaction mix (0.2 µl of FW primer, 0.2 µl of RW primer, 2.6 µl water, 5 µl of

SYBR-Green mix) with 2 µl of cDNA. The RT-PCR was conducted in ABI Prism® 7900HT Sequence Detection System following an initial denaturation step at 95°C for 5 minutes and then a fast cycle: 10 seconds 95°C, 60°C for 20 seconds repeated 40 times. The relative abundance of template cDNAs were normalised to GAPDH.

9.9.1 List of primers used

Gadph FW	TCTTGTGCAGTGCCAGCCT
Gadph RW	CAATATGGCCAAATCCGTTCA
Ctgf Intron FW	GCTCCTCGCTCTCTGCAC
Ctgf Intron RW	TGTGATCGCAGCTCACTCTG
Ctgf Exon FW	GGAGGAAAACATTAAGAAGGGCAA
Ctgf Exon RW	AACTTGACAGGCTTGGCGAT
Vcl Intron FW	CGTCACTTGCGTTGAGTACC
Vcl Intron RW	GAAACCACCCACAGGTTGGA
Vcl Exon FW	ACGGCTCTAGGGGAATCCTT
Vcl Exon RW	TTACGAACCTCAGCCTCATCG
Srf Intron FW	TCAAGGCAGCAGCAGTTTCT
Srf Intron RW	CAGGCAGGGTTAGGAACCAG
Srf Exon FW	TGAAGAAGGCCTATGAGCTGTC
Srf Exon RW	ACACATGGCCTGTCTCACTG
Acta2 Intron FW	CCAGAAGCAATGCGTCCACT
Acta2 Intron RW	TGAGGTAGTTGCCTGCTCTC
Acta2 Exon FW	CTGTCAGGAACCCTGAGACGC
Acta2 Exon RW	GGCTGTGCTGTCTTCCTCTT
Dusp5 Intron FW	GAGCACGTTAGGCTGTGTCT
Dusp5 Intron RW	GCCAGGGTCAAAAAGGCAAG
Dusp5 Exon FW	ACTGCTGACATTAGCTCCCAC
Dusp5 Exon RW	TTCTTCCCTGACACAGTCAATAAAA
Pdlim5 Intron FW	GCTTAGTTAGGATGGCCGCT
Pdlim5 Intron RW	CCTTGTGTCGACAGAGTGCT
Pdlim5 Exon FW	GCAAAATGGGAAAATTCCACCTA
Pdlim5 Exon RW	TGCGCTCCACAATGTGTTTT

Cyr61 Intron FW	CGTAAACTGCCCTGAGCCTA
Cyr61 Intron RW	GACGCGATCGAGACACTTCT
Cyr61 Exon FW	ATCGCAATTGGAAAAGGCAGC
Cyr61 Exon RW	GGTGCCAAAGACAGGAAGCCT
Serpine1 Intron FW	GTTGGAAGTGCGGTTTGACC
Serpine1 Intron RW	GCCTTTAGTCCACCCTAGCC
Serpine1 Exon FW	CATGTTTAGTGCAACCCTGGC
Serpine1 Exon RW	CTGCTCTTGGTCGGAAAGACT
Tpm1 Intron FW	AGGTGGGCCAGGATTCAAAC
Tpm1 Intron RW	CTGATCCGTGCCTGGCTAAC
Tpm1 Exon FW	GCCCGTAAGCTGGTCATCAT
Tpm1 Exon RW	CGGCACATTTGCCTTCTGAG
Thbs1 Intron FW	CTCCTTTATCCTGCCCCGTC
Thbs1 Intron RW	GTGAACCCGAAGGCTGAAGA
Thbs1 Exon FW	GAACAACGAGGAGTGGACTGT
Thbs1 Exon RW	TAACCGAGTTCTGGCAGTGAC
Palld Intron FW	CACAGCGCCATTGTCTTAGC
Palld Intron RW	GTTGCCGCAGAATGCCTTAG
Palld Exon FW	ACGTTCAGATCCACTGCGAG
Palld Exon RW	CGACCTGTGTCGTCCTCAAA
Bok Intron FW	GAACCTATCGCCACTTCCCT
Bok Intron RW	TAAATCTCGCAGATCCCAGCG
Bok Exon FW	CTTCTCAGCAGGTATCACATGG
Bok Exon RW	CCACGGAATACAGGGACACT
Ankrd1 Intron FW	CCACTCAACCCACTCAACCA
Ankrd1 Intron RW	GGAATAATGTGGGGTTGGCG
Ankrd1 Exon FW	GCCAGTTCCAGGGGTTCATC
Ankrd1 Exon RW	TTTTGCCTGTTACCAGCTCCT
Wdr1 Intron FW	CTGAGCACGGGAGAGACAAG
Wdr1 Intron RW	AATGAGTCAGTCTGCTGGGC
Wdr1 Exon FW	TGAGTACCAGCCTTTCGCTG
Wdr1 Exon RW	GCTCCAAACTTCTCCCTTCCT
D4Bwg0951e Intron FW	CTCGCCCTTGATGTGCTTGT
D4Bwg0951e Intron RW	GGGTAGGAGGCAGTCATTAGC

D4Bwg0951e Exon FW	AGGGAGCTACCTGGACACTT
D4Bwg0951e Exon RW	TCTGAAGGGGTCTGATGACCT
Pdlim7 Intron FW	TAGGGGCTCGGGATATGAGG
Pdlim7 Intron RW	AAAATGGCTAGGACCCGCTC
Pdlim7 Exon FW	TGTGGATCCTGCATTTGCTGA
Pdlim7 Exon RW	CACTGTGCTGGTTTTGTCTGG
Dstn Intron FW	ATAGCAACTGGCTTGCAGGT
Dstn Intron RW	ACAGCTAAGCATGGTCCGTT
Dstn Exon FW	CGAACATGGCCTCAGGAGTT
Dstn Exon RW	GGTGTGGAACATTTCCGAACTT
Vgll3 Intron FW	AGGTAGCAGAGGGTACCTGAG
Vgll3 Intron RW	CATCCGACCAGTGTTCCACC
Vgll3 Exon FW	GCCAGCAGAAGTTAGCGGTAT
Vgll3 Exon RW	TTTGCTGGGAAGCGTGACTT
Slc2a1 Intron FW	CCGGATTTACGGAACCCCTC
Slc2a1 Intron RW	GCAAAGGCGGGACAAGAAAG
Slc2a1 Exon FW	ATCTTCGAGAAGGCAGGTGTG
Slc2a1 Exon RW	CAACAAACAGCGACACCACAG
Klf6 Intron FW	TTGCGGAACGCAGTAGTTCT
Klf6 Intron RW	AACAAGGCAGGCATACGTGA
Klf6 Exon FW	ACGAAACGGGCTACTTCTCG
Klf6 Exon RW	CAGGCAGGTCTGTTGCCAAT
Rheb Intron FW	GAGAACGCAGAATGAACCGC
Rheb Intron RW	GGCAGTAGAACACCTTCCCC
Rheb Exon FW	GCTACCGGTCTGTGGGAAAG
Rheb Exon RW	TGGTGAACGTGTTCTCTATGGTT
Tpm4 Intron FW	AGTAGTAGGGCGGTGTTTGC
Tpm4 Intron RW	ACTACGGCTTCAGGAGACCT
Tpm4 Exon FW	CAAGTATGAGGAGGTTGCTCGT
Tpm4 Exon RW	TCAGATACCTCCGCCCTCTC
Klf7 Intron FW	CACTGGCTCCCTATACCGTG
Klf7 Intron RW	GATCCAAAGCAGGGTTTGCC
Klf7 Exon FW	CTCACACAGGTGAGAAGCCTTA
Klf7 Exon RW	CTTGTGAGCTCATCGCTCCG

Zyx Intron FW	CAACCTGGCTCGTTCTCACT
Zyx Intron RW	GACCATAACGAGGGGCTCAG
Zyx Exon FW	GCTACACCGACACTTTGGAGA
Zyx Exon RW	GTGAAGCACTGTGGGTGGTA
Pdcl3 Intron FW	GGCGGTCACTTCTCAAGGAA
Pdcl3 Intron RW	TGCTGCTGTCGACTGCTTAT
Pdcl3 Exon FW	AAGTTACGAAAGCCGGCGA
Pdcl3 Exon RW	AGAGGGGAATCCCTTGTTTGT
Actb Intron FW	CGTAGCGTCTGGTTCCCAAT
Actb Intron RW	GTGTGGGCATTTGATGAGCC
Actb Exon FW	CGCCACCAGTTCGCCAT
Actb Exon RW	CTTTGCACATGCCGGAGC
Egr1 Intron FW	TGATGTCTCCGCTGCAGATC
Egr1 Intron RW	GGTGGGTGAGTGAGGAAAGG
Egr1 Exon FW	ATTGATGTCTCCGCTGCAGATC
Egr1 Exon RW	TCAGCAGCATCATCTCCTCCA
Fos Intron FW	GCATGGGCTCTCCTGTCAA
Fos Intron RW	GACCTGGCGGCTACACAAA
Fos Exon FW	TTCCTACTACCATTCCCCAGCC
Fos Exon RW	GATCTGCGCAAAAGTCCTGTG
Egr2 FW	GAGCAAATGATGACCGCCAA
Egr2 RW	TGTCAGGCAGCTGGTGCATAA
Ier2 FW	CCTGCGGTTCCTTTGTCCTTA
Ier2 RW	TCACTTTGGTTTCCGACATGC
Nr4a1 FW	CTTCTGTCACCCATGTGCCTTT
Nr4a1 RW	TCAGTCTTAGCTCAAAACCAGGCT

9.10 RNA-seq library preparation and data collection

RNA-seq libraries were prepared by the Advanced Sequencing facility using directional mRNA-Seq Library Prep v1.0 Protocol from Illumina with minor adjustments. The samples were collected using GenElute Mammalian Total RNA Miniprep Kit and DNA contaminants were removed by treating the samples with DNaseI (see section above). To reduce the ribosomal rRNA in the libraries,

samples were processed using the Ribo-zero rRNA removal kit (Epicentre; separate CD-stimulation experiment). The Kapa HiFi HotStart ready mix, which reduced the overall volume of the PCR, was substituted to the Illumina kit Phusion enzyme and the ratio for the Agencourt AMPure XP beads was adjusted accordingly. The standard PCR cycle suggested by the Illumina protocol was changed to match the quantity of the total RNA measured initially with the Bioanalyser. The libraries were subject to cluster formation and then 72 single end sequencing using a Hiseq analyser.

The Bioinformatics and Biostatistics group performed the annotation and collection of RNA-seq reads. The RNA-seq data collected were all aligned to the NCBI37/mm9 with the BWA (version 0.5.6) using default settings. The reads were aligned to the canonical gene features annotated in RefSeq. All Reads and reads with sequences included in intronic features, were annotated using coverage Bed (bedtools version 2.14.3) using the bam file obtained by bwa as input. The data was then normalised as in (Esnault et al. 2014) using the same set of 6664 genes not changing across conditions. As described in the result section the differential gene expression analysis was performed using DESeq (Anders & Huber 2010) at $p<0.2$ for genes induced by Cyrochalasin D while genes induced by TPA in MEFs were considered at $p<0.05$.

The data were annotated in text files then analysed using Microsoft Excel and GraphPad Prism (GraphPad Software, Inc.).

9.11 Chromatin immunoprecipitation (ChIP)

ChIP assay was performed as previously described (Miralles et al. 2003) with few modifications. Each replica point was obtained by growing 3 15cm dishes of cells 60-80% confluent. Cells were fixed for 10 minutes at 37°C by supplementing the culturing media with a 4% formaldehyde solution in PBS to reach a final concentration of 1% formaldehyde in solution. The reaction was then stopped by adding glycine directly to the plates to reach a concentration of 250mM in solution. Each dish of cell was washed twice with 20ml of cold PBS. Cells were kept on ice using metallic plates and ice buckets. Cells were collected by scraping the dishes with a rubber policeman with 3ml of cold PBS supplemented with Protease inhibitors. At this stage the three plates were combined into a 15ml falcon

tube and pelleted at 3000rpm at 4°C for 5 minutes using a Beckman GS-6KR centrifuge. The pellet of cells was washed once in 5ml of Buffer A (Hepes pH8.0 5mM, KCl 85mM, 0.5% Triton-X-100, Protease Inhibitor cocktail, PMSF 1mM) and once with in 5ml of Buffer A' (Hepes pH8.0 5mM, KCl 85mM, Protease Inhibitor cocktail, PMSF 1mM). Finally the crude nuclei were resuspended in 1ml of Buffer B (Tris-HCl pH8 50mM, 1% SDS, 10 Mm EDTA, Protease Inhibitor cocktail, PMSF 1mM) and snap frozen on dry ice. At this stage cells could be kept at -80°C or thaw on ice to proceed with the sonication stage. DNA was sheared to reach an average of 300-500bp fragments using a Bioruptor® Plus (Diagenod) in 15ml Polystyrene tubes (blue cap) with automatic water bath cooling system. Before sonicating I checked that all the SDS contained in the sample was homogenously dissolved as this can cause changes in the fragment sizes. Four cycles of 30 s low intensity sonication followed by 30 s of pause were performed to reach the desired fragment size. Samples were sonicated three by three using aluminium rings to fit the falcon in the sonicator wheel The supernatant was collected by centrifugation and diluted 1 in 10 in FA/SDS like buffer (Hepes KOH pH 7.5 50mM, NaCl 150mM, Triton-X-100 1%, Na deoxycholate 0.1%, Protease Inhibitor cocktail, PMSF 1mM). 0.1ml of the diluted sample were kept as input sample, reverse-crosslinked overnight and purified using the QIAquick PCR Purification Kit.

Each IP was performed with 1ml of the diluted sample (with 3 dishes we performed up to 10 IP with different antibodies). 0.5mg of magnetic Dynabeads (either IgG or IgM according to the type of antibody used) was used for each IP. The beads were washed three times with 1ml of PBS-BSA solution (PBS, 0.1% BSA w/v) and incubated with the desired antibody at 4°C under agitation for 1 hour. In parallel 0.25mg of beads were also washed three times in PBS-BSA and mixed with the 1ml sample to pre-clear unspecific bindings for 1 hour in the cold under agitation. The beads coupled to the antibody were then washed in clean tubes 5 times with PBS-BSA. The washed beads coupled to the antibody were combined with the pre-cleared sample and left over night at 4°C under agitation. The day after the samples were collected, transferred to clean tubes and washed three times with 1ml of FA/SDS solution (Hepes KOH pH 7.5 50mM, NaCl 150mM, Triton-X-100 1%, Na deoxycholate 0.1%, EDTA 1mM, SDS 0.1%, Protease Inhibitor cocktail, PMSF 1mM). At the last wash the samples were incubated for 10 minutes at 4°C under agitation. Finally the samples were washed with 1ml of WB

Buffer (Tris-HCl pH 8 10mM, LiCl 0.25M, EDTA 1mM, NP40 0.5%, Na deoxycholate 0.5%) and with 1ml of TE Buffer (Tris-HCl pH 8 10mM, EDTA 1mM). The Immune complexes were then eluted with 125µl of Elution Buffer (Tris-HCl pH 7.5 25mM, EDTA 5mM, SDS 0.5%) at 65°C for 25 minutes using and Eppendorf thermocycler under agitation. The samples were then reverse cross linked over night with 10µl of Proteinase K at 65°C. The samples were then collected and purified using the QIAquick PCR Purification Kit.

The samples were analysed using RT-PCR as described in the previous sections. The same samples were used to prepare ChIP-seq libraries.

9.11.1 List of antibodies used

Antibodies used were: H3K4me3 (AB1012, Abcam); H3K9ACS10P (AB12181X510/AC, Abcam); H4K16ac (AB61240, Abcam); Histone H3 (AB1791, Abcam); H3K27ac (AB4729, Abcam); H3K27me3 (AB6002, Abcam); Thr4P Pol II (6D7, ActiveMotif); Ser7P PolII (4E12, ActiveMotif); MED1 (A300793A, Bethnyl Lab); Ser5P PolII (H14, Covance); Ser2P PolII (H5, Covance); Tyr1P PolII (3D12, kindly provided by Kindly provided by Dirk Eick,Helmholtz-Zentrum Muenchen); H3K9K14ac (06-599, Millipore); H3S10P (04-817, Millipore), H3S0P (C15410116/pAb-116-050, Diagenode); SRF (sc-335, S. Cruz); PolII N20(sc-899, S. Cruz); MRTF-A (sc-21558, S. Cruz); NELF-A (sc-23599, S. Cruz); SPT5 (sc-28678, S. Cruz); CDK9 (sc-484, S. Cruz); Pol II 8WG16 (sc-56767, S. Cruz); HA (1867431, Roche).

The Elk-1 antibody was obtained from an in-house affinity-purified anti-mouse Elk-1 aa309–429 (Costello et al. 2010; Buchwalter et al. 2005). The antibody was re-purified against a peptide derived from the human Elk-1 NONA proteins (aa309-429).

9.11.2 List of primers used for ChIP RT-PCR

ACTA2 UP f	CCTGCAAGCCAAGGTTCTGA
ACTA2 UP r	GCACTCCCAGAATCCATCCA
ATCA2 SRE(-0.3Kb) FW	GAGGCCTGGGTCTCTTCCA
ATCA2 SRE(-0.3Kb) RW	GCTGAGCTGCCTCCTGTTTC
ACTA2 TSS(0Kb) FW	CATTCAGATTCCCACAGACAATG

ACTA2 TSS(0Kb) RW	TCGAGTTTTCCCAGGCTCTTT
ACTA2 (2.8Kb) FW	CCCACGATGGATGGGAAA
ACTA2 (2.8Kb) RW	CGGCTTCGCTGGTGATG
ACTA2 (6.6Kb) FW	AGAGGAGCGTGGTAATCTGTTCTT
ACTA2 (6.6Kb) FW	GAACATCCCTGTCCCTTTCCA
ACTA2 (12.3Kb) FW	ACTGGACCCCTGAGTTTCACA
ACTA2 (12.3Kb) RW	GGCAAGCCTCAATTCTCCAA
ACTA2 (14.1Kb) FW	ATACGCAAGGCTTGATGCAA
ACTA2 (14.1Kb) RW	AAGCATACACACGTGCATGGA
ACTA2 (16.3Kb) FW	CCGTATTTGAATCTGCAACATTCT
ACTA2 (16.3Kb) RW	TATAACACAAGAGCAAATGGCTGAA
Actb FW TSS 0kb	AGGAGCTGCAAAGAAGCTGT
Actb RW TSS 0Kb	CCGCTGTGGCGTCCTATAAA
Actb FW 1Kb	GGCTTTGCACATGCCGGA
Actb RW 1Kb	CTTTTGTGTCTTGATAGTTCGCCA
Actb FW 2Kb	CGGAGTCCATCACAATGCCT
Actb RW 2Kb	GCCATGTACGTAGCCATCCA
Actb FW TES 3.1Kb	GCCTTCACCGTTCCAGTTTT
Actb RW TES 3.1Kb	TGAGCTGCGTTTTACACCCT
Actb FW 4.1Kb out	GTCCAAGGATCACGACTGACA
Actb RW 4.1Kb out	CATCCTGGAAATCAGCCCCT
Zfp37 FW	CCAGCAATGTGTGACTTGGATC
Zfp37 RW	TATTTCGAGCGCTGTGGCA
SRF FW	ATACCGAACTCGCTGCTGTCAT
SRF RW	AACTGGTTCGGCTCCACTGTT
Cyr61 FW	ATGCCTTGTGGTTGGATAACAGAGG
Cyr61 RW	CCAGATGGTGAATCAGACACCAGAC
Klf7 FW	GGAGGAGGGCGTCCATTAG
Klf7 RW	CATGAGCCCCCTGTTTACCTT
Col1a1 FW	CCAGGAGGGCATATGGAAGA
Col1a1 RW	GTCCTCAGCCCCTTATTTGGT
Actg1 FW	AACGCGGTGCACGAGAAG
Actg1 RW	TCACACTGCCCAGTTGCAA
Myh9 FW	CCTGGTGCTACCATAAAAGGAAA

Myh9 RW	GGACAGCCCTGGGAAACAGT
Mir145-143 FW	CCTTGCCCGTGGCTCTCT
Mir145-143 RW	AGGCTCGTTTCTTCAGCTCATATAA
Actb FW	GCCGCCGGGTTTTATAGG
Actb RW	CGTTCCGAAAGTTGCCTTTTA
FilaminA FW	TGAGCTCAGCGCTCTGTGAA
FilaminA RW	GCTCTGGAGGTGAGCCCTACT
EGR1-1 F	TGGAAACAAGAGCCTCCCATT
EGR1-1 R	GAAGCCCTATCTCCGAAGCA
EGR1-2 F	GCCATATAAGGAGCAGGAAGGA
EGR1-2 R	AAGGCGCTGCCCAAATAAG
EGR1-3 F	TGATGTCTCCGCTGCAGATC
EGR1-3 R	GGTGGGTGAGTGAGGAAAGG
EGR1-4 F	AAGCCTTTTGCCTGTGACATTT
EGR1-4 R	ATGCCTCTTGCGTTCATCACT

9.12 ChIP-seq library preparation and data collection

ChIP-seq libraries were prepared by the Advanced Sequencing facility. The DNA samples were end repaired, poly-A tailed and the Illumina signle end adapters were ligated following the protocol provided by Illumina. Agencourt AMPure XP beads were used to remove the adaptors. As for the RNA-seq library the Kapa HiFi HotStart ready mix was substituted to the Illumina kit Phusion enzyme and the ratio for the Agencourt AMPure XP beads was adjusted accordingly. 15 cycle of PCR were performed to amplify the library and the library was size selected performing agarose gel electrophoresis (2% Agarose gel, NB E-gels were avoided as we estimated a consistent loss of material during the process of extraction). Libraries with sizes between 200bp and 1Kb were cut from the gel and extracted using the MinElute Gel Extraction Kit (Qiagen). As described previously the purified library was then quality controlled on the DNA 1000 BioAnalyser 2100 chip. Clusters were assembled and subsequent 36-50bp single end sequencing on a Hi-seq 2000/1000.

The ChIP-seq reads were aligned using Eland (version pipeline 1.4) to the to the NCBI37/mm9 mouse genome using the default settings. The frequency

distribution of the reads was analysed to assess homogenous distribution and discharge samples affected by over-amplification.

9.12.1 ChIP-seq SRF and MRTF peak calling

MACS version 1.4.2 was used to identify regions enriched over background (IP using beads only as negative control). MACS threshold of $p<10^{-5}$ were used. The ChIP-seq mapped reads were normalised to 30 milions total reads and the read density per base was calculated for each enriched region. As described in the result section SRF and MRTF peaks were crossed with the set of SRF peaks already described (Esnault et al. 2014). Peaks were considered as equal if the coordinates per peak defined by MACS were overlapping for at least 1 bp. Given the size of the genome, while 1bp overlap maybe arbitrary, if 2 peaks overlap then we can say that is a true overlap, as the search space is large. We considered defining the standard deviation of each bell shape peak and define the overlap between peaks as fraction of this number but the shape of the peaks is not normally distributed and as a result it would have been less statistically valid.

9.12.2 RNA Pol II analysis and normalisation

Pol II, Ser5P and Ser2P ChIP-seq data were aligned to the mouse genome, version NCBI37/mm9 using Eland (version 1.4). The mapped reads were normalised to 30 millions comparing the distribution of reads mapped in gene feature of constitutively expressed genes (invariant genes as assessed by RNA-seq and analysed by DeSeq showing at least 5 reads in intronic features). A scaling factor was obtained by generating Bland-Altman plots comparing each condition to the 0.3% FCS for the invariant set of genes. As constitutively expressed genes should show equal amount of Pol II signal within their gene features, the distribution of the differences across conditions should assume a Gaussian distribution. The mean difference between the distributions obtained reflects the systematic error across the sample under analysis. Therefore each Pol II plot was scaled to the reference set of invariant constitutively expressed genes.

For all UCSC mm9 annotated gene we quantified the signal counting the reads in 250bp windows from -2Kb from the TSS to the annotated polyA site.

These values were used to generate scatter and box plots to compare the amount of Pol II signal per groups of genes.

9.12.3 Density plots for SRF, MRTF and Pol II

The nucleotide average read density was used to generate density plots for SRF, MRTF-A and Pol II. In the case of SRF and MRTF-A, the average read density was calculated by centring on the SRF or MRTF-A ChIP-seq binding loci (±2 kb around the SRF or MRTF-A ChIP-seq signal summit).

For the Pol II data sets the density profiles were calculated across the gene loci ±5 kb. The density plots were then scaled according to the invariant set as described in the previous section.

9.13 Go analysis

As described in the result section Gene Ontology analysis were performed using DAVID (Huang et al. 2009). Functional terms were collected per group of genes using the Functional annotation clustering, obtained with David, which group similar annotations together to reduce the redundancy of the annotations. One term per cluster, with lowest p-value and higher number of gene included, was selected per functional cluster. The same term was then search into the Functional annotation chart, which identify enriched annotation terms associated to the list examined, in order to identify the enrichment p-value associated to the functional term compared to the list of genes analysed. The Functional annotation chart is based on an enrichment analysis therefore is highly redundant. The selection of unique terms using the DAVID functional clustering avoids the selection of redundant categories. The p-value displayed for each term was collected from the functional chart in order to express the enrichment of that category within the list of gene provided. Therefore is not possible to compare the enrichment of functional categories across different gene groups. However the data displayed provides information regarding how well one group is enriched compared to another within the same gene list.

9.14 Nuclear Run-On assay

The nuclear Run-On protocol was performed as previously described by others (Preker et al. 2008; Kwak et al. 2013; Hah et al. 2011). Each replicate derived from one 15cm dish ($6X10^6$ cells 80% confluent) treated as described in the result section. Each dish of cell was washed 3 times with cold PBS and placed on ice over metallic plates. 10ml of Swelling Buffer (10mM TRis pH 7.5, 2mM $MgCl_2$, 3mM $CaCl_2$) were added to the cell layer directly in the dish and let to incubate on ice for 5 minutes. The cells were then scraped with rubber policeman, transferred to a 15ml falcon tube and centrifuged in a Beckman GS-GR tabletop centrifuge at 1200rpm for 5 minutes at 4°C. After removing the supernatant 1ml of Lysis Buffer (10mM TRis pH 7.5, 2mM $MgCl_2$, 3mM $CaCl_2$, 0.5% NP40, 10%glycerol, 2Units/mL SuperaseIn from Ambion) was add to the pellet resuspended seveal times using a p1000 pipet and let to incubate on ice for 10 minutes. The lysate was diluted 10 times using the Lysis Buffer and pelleted at 2000 rpm in Beckman GS-GR tabletop centrifuge. The pellet was washed again in 10ml of Lysis Buffer and pelleted again at 2000 rpm in Beckman GS-GR tabletop centrifuge. The Nucleai pellet was then washed once in 1ml of Freez buffer (50mM Hepes pH 8.2, 40%Glycerol, 5mM $MgCl_2$, 0.1mMEDTA), centrifuged at 1000 xg and resuspended in 0.1ml and snap frozen in dry ice and kept at -80°C.

100µl of nuclei were thaw on ice and mixed with 100µl of Reaction Buffer (10 mM Tris–HCl pH 8.0, 5 mM $MgCl_2$, 1 mM DTT, 300 mM KCl, 20 U of RNase inhibitor, 1% sarkosyl, 0.5 mM of BrUTP, ATP, and GTP, and 5 mM CTP) was add to carry out the Run-on assay for 5 minutes at 30°C ina thermo shaker. The reaction was stop by adding 0.5ml of Trizol (Sigma) to the mix and the RNA extracted according to the manufacturer's protocol. The RNA was extracted and precipitated using isopropanol and the pellet resuspended in 20µl of DEPC water. The Run-on reaction could also be stop by adding 50 U DNase I and Proteinase K at 37°C for 1h. The RNA was subject to base hydrolysis by adding 5µl of 1M NaOH solution and incubating on ice for 40 minutes. The hydrolysis reaction was stop by adding 25µl of 1M Tris-HCl pH 6.8. The RNA was then extracted using GenElute Mammalian Total RNA Miniprep Kit following the manufacturer instruction (Sigma) and eluted in 70µl. The sample was treated with 10µl of RQ1 DNaseI (Promega) and incubated for 10 minutes at 37°C with 6.7µl of RQ1 DNaseI buffer. The RNA

was then re-extracted as above using GenElute Mammalian Total RNA Miniprep Kit. 100µl of anti-BrdU agarose beads (Santa Cruz Biotech) were blocked in Blocking solution (0.5SSPE, 1mMEDTA, 0.05% Tween, 0.1%PolyVinylPirrolidone 1mg/ml BSA) for 1h at 4°C under rotation. 85µl of the treated RNA were mixed with the beads in 500µl of binding buffer (0.5XSSPE, 1mM EDTA, 0.05%Tween) and incubated under rotation for 1h at 4°C. The beads were then washed once in LowSalt Buffer (0.2X SSPE, 1mM EDTA, 0.05%Tween), twice in HighSalt Buffer (0.5%SSPE, 1mM EDTA, 0.05% Tween) and twice in TETBuffer (TE Ph 7.4, 0.05% Tween). The RNA was then eluted with 125µl of Buffer E (20mM DTT, 300mM KCl, 5mM Tris-HCl ph7.5, 1mM EDTA, 0.1% SDS), phenol-chloroform extracted and ethanol precipitated overnight. The collected RNA was then used to perform reverse transcription and RT-PCR as described above.

9.15 DIP/DRIP assay

The DNA-R-loop IP (DRIP) was performed using an antibody recognizing RNA/DNA hybrids purified from S9.6 hybridoma cell lines (Boguslawski et al. 1986) and kindly provided by Nicholas J Proudfoot. The protocol follows overall the step described in the ChIP protocol. Cells were not crosslinked. After Nuclear lysis the DNA was treated with 30mg of proteinase K (Roche) at 55°C for 3 hours. The DNA was then extracted by Ethanol precipitations, resuspendent in 1ml Buffer B (see ChIP protocol) and sonicated as in the ChIP protocol. The samples were diluted 10 times in Buffer FA/SDS like (see ChIP protocol) and the DIP analysis carried out using the S9.6 antibody. Each 0.5mg of dynabeads were mixed with 10µl of the purified S9.6 antibody and used with 1ml of the diluted DNA. The IP, supernatant and the input (collected as for the ChIP protocol, 100µl of the diluted DNA) were purified using the QIAquick PCR Purification Kit and used in Q-PCR. The same IP was performed using 1ml samples previously treated with 10 U of RNase H (NEB) for 2 hr at 37°C as control.

Reference List

Adachi, M., Fukuda, M. & Nishida, E., 1999. Two co-existing mechanisms for nuclear import of MAP kinase: passive diffusion of a monomer and active transport of a dimer. *The EMBO Journal*, 18(19), pp.5347–5358.

Adams, C.C. & Workman, J.L., 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Molecular and cellular biology*, 15(3), pp.1405–1421.

Adcock, I.M., 2006. Histone deacetylase inhibitors as novel anti-inflammatory agents. *Current opinion in investigational drugs (London, England : 2000)*, 7(11), pp.966–973.

Adelman, K. & Lis, J.T., 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews Genetics*, 13(10), pp.720–731.

Aguilera, A. & García-Muse, T., 2012. R loops: from transcription byproducts to threats to genome stability. *Molecular Cell*, 46(2), pp.115–124.

Akhtar, M.S. et al., 2009. TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Molecular Cell*, 34(3), pp.387–393.

Akoulitchev, S., Chuikov, S. & Reinberg, D., 2000. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature*, 407(6800), pp.102–106.

Amoutzias, G.D. et al., 2008. Choose your partners: dimerization in eukaryotic transcription factors. *Trends in Biochemical Sciences*, 33(5), pp.220–229.

Anders, S. & Huber, W., 2010. Differential expression analysis for sequence count data. *Genome biology*, 11(10), p.R106.

Anderson, J.D. & Widom, J., 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Molecular and cellular biology*, 21(11), pp.3830–3839.

Aoki, K. et al., 2013. Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Molecular Cell*, 52(4), pp.529–540.

Appella, E. & Anderson, C.W., 2000. Signaling to p53: breaking the posttranslational modification code. *Pathologie-biologie*, 48(3), pp.227–245.

Aravind, L. & Koonin, E.V., 2000. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences*, 25(3), pp.112–114.

Archambault, J. et al., 1997. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 94(26), pp.14300–14305.

Baarlink, C., Wang, H. & Grosse, R., 2013. Nuclear actin network assembly by formins regulates the SRF coactivator MAL. *Science*, 340(6134), pp.864–867.

Balamotis, M.A. et al., 2009. Complexity in Transcription Control at the Activation Domain-Mediator Interface. *Science Signaling*, 2(69), pp.ra20–ra20.

Banerji, J., Rusconi, S. & Schaffner, W., 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *CELL*, 27(2 Pt 1), pp.299–308.

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell Research*, 21(3), pp.381–395.

Bao, Y. & Shen, X., 2007. SnapShot: Chromatin Remodeling Complexes. *CELL*, 129(3), pp.632.e1–632.e2.

Barboric, M. et al., 2001. NF-kappaB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Molecular Cell*, 8(2), pp.327–337.

Bardwell, L. et al., 2007. Mathematical Models of Specificity in Cell Signaling. *Biophysical journal*, 92(10), pp.3425–3441.

Barratt, M.J. et al., 1994. A mitogen- and anisomycin-stimulated kinase phosphorylates HMG-14 in its basic amino-terminal domain in vivo and on isolated mononucleosomes. *The EMBO Journal*, 13(19), pp.4524–4535.

Barski, A. et al., 2007. High-Resolution Profiling of Histone Methylations in the Human Genome. *CELL*, 129(4), pp.823–837.

Bartke, T. et al., 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *CELL*, 143(3), pp.470–484.

Bartkowiak, B. et al., 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & Development*, 24(20), pp.2303–2316.

Bartolomei, M.S. et al., 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Molecular and cellular biology*, 8(1), pp.330–339.

Basehoar, A.D., Zanton, S.J. & Pugh, B.F., 2004. Identification and distinct regulation of yeast TATA box-containing genes. *CELL*, 116(5), pp.699–709.

Baskaran, R. et al., 1997. Ataxia telangiectasia mutant protein activates c-Abl tyrosine kinase in response to ionizing radiation. *Nature*, 387(6632), pp.516–519.

Bataille, A.R. et al., 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Molecular Cell*, 45(2), pp.158–170.

Baumli, S. et al., 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *The EMBO Journal*, 27(13), pp.1907–1918.

Belchetz, P.E. et al., 1978. Hypophysial responses to continuous and intermittent delivery of hypopthalamic gonadotropin-releasing hormone. *Science*, 202(4368), pp.631–633.

Belotserkovskaya, R. et al., 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science*, 301(5636), pp.1090–1093.

Bengal, E. et al., 1991. Role of the mammalian transcription factors IIF, IIS, and IIX during elongation by RNA polymerase II. *Molecular and cellular biology*, 11(3), pp.1195–1206.

Bentley, D.L., 2014. Coupling mRNA processing withtranscription in time and space. *Nature Reviews Genetics*, 15(3), pp.163–175.

Bentley, D.L. & Groudine, M., 1986. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature*, 321(6071), pp.702–706.

Bernecky, C. & Taatjes, D.J., 2012. Activator-mediator binding stabilizes RNA polymerase II orientation within the human mediator-RNA polymerase II-TFIIF assembly. *Journal of molecular biology*, 417(5), pp.387–394.

Bernstein, B.E. et al., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *CELL*, 125(2), pp.315–326.

Bernstein, B.E., Meissner, A. & Lander, E.S., 2007. The mammalian epigenome. *CELL*, 128(4), pp.669–681.

Besnard, A. et al., 2011. Elk-1 a transcription factor with multiple facets in the brain. *Frontiers in neuroscience*, 5, p.35.

Boguslawski, S.J. et al., 1986. Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. *Journal of immunological methods*, 89(1), pp.123–130.

Bonasio, R., Tu, S. & Reinberg, D., 2010. Molecular signals of epigenetic states. *Science*, 330(6004), pp.612–616.

Borneman, A.R. et al., 2007. Divergence of transcription factor binding sites across related yeast species. *Science*, 317(5839), pp.815–819.

Bortvin, A. & Winston, F., 1996. Evidence that Spt6p controls chromatin structure by a

direct interaction with histones. *Science*, 272(5267), pp.1473–1476.
Bourbon, H.-M., 2008. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic Acids Research*, 36(12), pp.3993–4008.
Bourbon, H.-M. et al., 2004. A Unified Nomenclature for Protein Subunits of Mediator Complexes Linking Transcriptional Regulators to RNA Polymerase II. *Molecular Cell*, 14(5), pp.553–557.
Boyer, T.G. et al., 1999. Mammalian Srb/Mediator complex is targeted by adenovirus E1A protein. *Nature*, 399(6733), pp.276–279.
Boyes, J. & Felsenfeld, G., 1996. Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *The EMBO Journal*, 15(10), pp.2496–2507.
Bösken, C.A. et al., 2014. The structure and substrate specificity of human Cdk12/Cyclin K. *Nature communications*, 5, p.3505.
Brandeis, M. et al., 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature*, 371(6496), pp.435–438.
Brannan, K. et al., 2012. mRNA Decapping Factors and the Exonuclease Xrn2 Function in Widespread Premature Terminationof RNA Polymerase II Transcription. *Molecular Cell*, 46(3), pp.311–324.
Brent, R. & Ptashne, M., 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *CELL*, 43(3 Pt 2), pp.729–736.
Brookes, E. et al., 2012. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell stem cell*, 10(2), pp.157–170.
Brown, M.A. et al., 2006. Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Molecular cancer*, 5, p.26.
Buchwalter, G., Gross, C. & Wasylyk, B., 2004. Ets ternary complex transcription factors. *Gene*, 324, pp.1–14.
Buchwalter, G., Gross, C. & Wasylyk, B., 2005. The ternary complex factor Net regulates cell migration through inhibition of PAI-1 expression. *Molecular and cellular biology*, 25(24), pp.10853–10862.
Bugyi, B. & Carlier, M.-F., 2010. Control of actin filament treadmilling in cell motility. *Annual Review of Biophysics*, 39, pp.449–470.
Bulger, M. & Groudine, M., 2011. Functional and Mechanistic Diversity of Distal Transcription Enhancers. *CELL*, 144(3), pp.327–339.
Burke, T.W. & Kadonaga, J.T., 1997. The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. *Genes & Development*, 11(22), pp.3020–3031.
Butler, J.S. & Platt, T., 1988. RNA processing generates the mature 3' end of yeast CYC1 messenger RNA in vitro. *Science*, 242(4883), pp.1270–1274.
Cai, G. et al., 2009. Mediator structural conservation and implications for the regulation mechanism. *Structure (London, England : 1993)*, 17(4), pp.559–567.
Cai, H. et al., 2014. Nucleocytoplasmic Shuttling of a GATA Transcription Factor Functions as a Development Timer. *Science*, 343(6177), pp.1249531–1249531.
Calhoun, V.C., Stathopoulos, A. & Levine, M., 2002. Promoter-proximal tethering elements regulate enhancer-promoter specificity in the Drosophila Antennapedia complex. *Proceedings of the National Academy of Sciences*, 99(14), pp.9243–9247.
Calvo, F. et al., 2013. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. *Nature cell biology*, 15(6), pp.637–646.
Cantin, G.T., Stevens, J.L. & Berk, A.J., 2003. Activation domain–mediator interactions

promote transcription preinitiation complex assembly on promoter DNA. *Proceedings of the National Academy of Sciences*, 100(21), pp.12003–12008.
Carninci, P. et al., 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature genetics*, 38(6), pp.626–635.
Carninci, P. et al., 2005. The transcriptional landscape of the mammalian genome. *Science*, 309(5740), pp.1559–1563.
Carr, A. & Biggin, M.D., 1999. A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. *The EMBO Journal*, 18(6), pp.1598–1608.
Chambon, P., 1975. Eukaryotic nuclear RNA polymerases. *Annual Review of Biochemistry*, 44, pp.613–638.
Chang, J.H. et al., 2012. Dxo1 is a new type of eukaryotic enzyme with both decapping and 5“-3” exoribonuclease activity. *Nature Structural & Molecular Biology*, 19(10), pp.1011–1017.
Chapman, R.D. et al., 2008. Molecular evolution of the RNA polymerase II CTD. *Trends in genetics : TIG*, 24(6), pp.289–296.
Chen, H.-T. & Hahn, S., 2004. Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *CELL*, 119(2), pp.169–180.
Chen, T. & Dent, S.Y.R., 2014. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature Reviews Genetics*, 15(2), pp.93–106.
Chen, X. et al., 2008. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *CELL*, 133(6), pp.1106–1117.
Chen, X.-F. et al., 2012. Mediator and SAGA have distinct roles in Pol II preinitiation complex assembly and function. *Cell reports*, 2(5), pp.1061–1067.
Chen, Y. et al., 2009. DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation. *Genes & Development*, 23(23), pp.2765–2777.
Chesnut, J.D., Stephens, J.H. & Dahmus, M.E., 1992. The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. *The Journal of biological chemistry*, 267(15), pp.10500–10506.
Chiu, Y.-L. et al., 2002. Tat stimulates cotranscriptional capping of HIV mRNA. *Molecular Cell*, 10(3), pp.585–597.
Cho, H. et al., 1999. A protein phosphatase functions to recycle RNA polymerase II. *Genes & Development*, 13(12), pp.1540–1552.
Cho, Y.-W. et al., 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *The Journal of biological chemistry*, 282(28), pp.20395–20406.
Chu, W.-Y. et al., 2009. ProteDNA: a sequence-based predictor of sequence-specific DNA-binding residues in transcription factors. *Nucleic Acids Research*, 37, pp.W396–401.
Cieślik, M. & Bekiranov, S., 2014. Combinatorial epigenetic patterns as quantitative predictors of chromatin biology. *BMC genomics*, 15, p.76.
Cirillo, L.A. et al., 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *The EMBO Journal*, 17(1), pp.244–254.
Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2), pp.279–289.
Clapier, C.R. & Cairns, B.R., 2009. The Biology of Chromatin Remodeling Complexes. *Annual Review of Biochemistry*, 78(1), pp.273–304.
Clark, K.L. et al., 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature*, 364(6436), pp.412–420.
Clark, T.G. & Merriam, R.W., 1977. Diffusible and bound actin nuclei of Xenopus laevis

oocytes. *CELL*, 12(4), pp.883–891.
Clayton, A.L., Hazzalin, C.A. & Mahadevan, L.C., 2006. Enhanced Histone Acetylation and Transcription: A Dynamic Perspective. *Molecular Cell*, 23(3), pp.289–296.
Conaway, R.C. & Conaway, J.W., 1990. Transcription initiated by RNA polymerase II and purified transcription factors from liver. Transcription factors alpha, beta gamma, and delta promote formation of intermediates in assembly of the functional preinitiation complex. *The Journal of biological chemistry*, 265(13), pp.7559–7563.
Conaway, R.C., Bradsher, J.N. & Conaway, J.W., 1992. Mechanism of assembly of the RNA polymerase II preinitiation complex. Evidence for a functional interaction between the carboxyl-terminal domain of the largest subunit of RNA polymerase II and a high molecular mass form of the TATA factor. *Journal of Biological Chemistry*, 267(12), pp.8464–8467.
Connelly, S. & Manley, J.L., 1988. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & Development*, 2(4), pp.440–452.
Core, L.J. & Lis, J.T., 2008. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*, 319(5871), pp.1791–1792.
Costello, P. et al., 2010. Ternary complex factors SAP-1 and Elk-1, but not net, are functionally equivalent in thymocyte development. *Journal of immunology (Baltimore, Md. : 1950)*, 185(2), pp.1082–1092.
Courey, A.J. & Tjian, R., 1988. Analysis of Sp1 in vivo reveals mutiple transcriptional domains, including a novel glutamine-rich activation motif. *CELL*, 55(5), pp.887–898.
Cowling, V.H., 2010. Myc up-regulates formation of the mRNA methyl cap. *Biochemical Society transactions*, 38(6), pp.1598–1601.
Cowling, V.H. & Cole, M.D., 2010. Myc Regulation of mRNA Cap Methylation. *Genes & Cancer*, 1(6), pp.576–579.
Cramer, P., 2001. Structural Basis of Transcription: RNA Polymerase II at 2.8 Angstrom Resolution. *Science*, 292(5523), pp.1863–1876.
Cramer, P. et al., 2008. Structure of Eukaryotic RNA Polymerases. *Annual Review of Biophysics*, 37(1), pp.337–352.
Creemers, E.E. et al., 2006. Coactivation of MEF2 by the SAP domain proteins myocardin and MASTR. *Molecular Cell*, 23(1), pp.83–96.
Cress, W.D. & Triezenberg, S.J., 1991. Critical structural elements of the VP16 transcriptional activation domain. *Science*, 251(4989), pp.87–90.
Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50), pp.21931–21936.
Cruzalegui, F.H., Cano, E. & Treisman, R., 1999. ERK activation induces phosphorylation of Elk-1 at multiple S/T-P motifs to high stoichiometry. *Oncogene*, 18(56), pp.7948–7957.
Cuevas, I.C., Frasch, A.C.C. & D'Orso, I., 2005. Insights into a CRM1-mediated RNA-nuclear export pathway in Trypanosoma cruzi. *Molecular and biochemical parasitology*, 139(1), pp.15–24.
Curran, T. et al., 1982. FBJ murine osteosarcoma virus: identification and molecular cloning of biologically active proviral DNA. *Journal of virology*, 44(2), pp.674–682.
Czudnochowski, N., Bösken, C.A. & Geyer, M., 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature communications*, 3, p.842.
Dalton, S. & Treisman, R., 1992. Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. *CELL*, 68(3), pp.597–612.
Danko, C.G. et al., 2013. Signaling pathways differentially affect RNA polymerase II

initiation, pausing, and elongation rate in cells. *Molecular Cell*, 50(2), pp.212–222.
Darimont, B.D. et al., 1998. Structure and specificity of nuclear receptor-coactivator interactions. *Genes & Development*, 12(21), pp.3343–3356.
David, C.J. & Manley, J.L., 2011. The RNA polymerase C-terminal domain: a new role in spliceosome assembly. *Transcription*, 2(5), pp.221–225.
Davis, R.L., Weintraub, H. & Lassar, A.B., 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *CELL*, 51(6), pp.987–1000.
de Ruijter, A.J.M. et al., 2003. Histone deacetylases (HDACs): characterization of the classical HDAC family. *The Biochemical journal*, 370(Pt 3), pp.737–749.
DerMardirossian, C. & Bokoch, G.M., 2005. GDIs: central regulatory molecules in Rho GTPase activation. *Trends in cell biology*, 15(7), pp.356–363.
Deschamps, J., Meijlink, F. & Verma, I.M., 1985. Identification of a transcriptional enhancer element upstream from the proto-oncogene fos. *Science*, 230(4730), pp.1174–1177.
Devaiah, B.N. et al., 2012. BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proceedings of the National Academy of Sciences of the United States of America*, 109(18), pp.6927–6932.
Dichtl, B. et al., 2002. A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. *Molecular Cell*, 10(5), pp.1139–1150.
Dominguez, R. & Holmes, K.C., 2011. Actin structure and function. *Annual Review of Biophysics*, 40, p.169.
Dong, C. et al., 2000. Microtubule binding to Smads may regulate TGF beta activity. *Molecular Cell*, 5(1), pp.27–34.
Donner, A.J. et al., 2010. CDK8 is a positive regulator of transcriptional elongation within the serum response network. *Nature Structural & Molecular Biology*, 17(2), pp.194–201.
Dopie, J. et al., 2012. Active maintenance of nuclear actin by importin 9 supports transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), pp.E544–52.
Drobic, B. et al., 2010. Promoter chromatin remodeling of immediate-early genes is mediated through H3 phosphorylation at either serine 28 or 10 by the MSK1 multi-protein complex. *Nucleic Acids Research*, 38(10), pp.3196–3208.
Drysdale, C.M. et al., 1995. The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids. *Molecular and cellular biology*, 15(3), pp.1220–1233.
Du, K.L. et al., 2004. Megakaryoblastic leukemia factor-1 transduces cytoskeletal signals and induces smooth muscle cell differentiation from undifferentiated embryonic stem cells. *The Journal of biological chemistry*, 279(17), pp.17578–17586.
Dunham, I. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 488(7414), pp.57–74.
Dupont, S. et al., 2011. Role of YAP/TAZ in mechanotransduction. *Nature*, 474(7350), pp.179–183.
Dye, M.J. & Proudfoot, N.J., 2001. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *CELL*, 105(5), pp.669–681.
Eberhardy, S.R. & Farnham, P.J., 2002. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *The Journal of biological chemistry*, 277(42), pp.40156–40162.
Ebinu, J.O. et al., 1998. RasGRP, a Ras guanyl nucleotide- releasing protein with calcium- and diacylglycerol-binding motifs. *Science*, 280(5366), pp.1082–1086.
Egly, J.-M. & Coin, F., 2011. A history of TFIIH: two decades of molecular biology on a pivotal transcription/repair factor. *DNA repair*, 10(7), pp.714–721.
Egyházi, E. et al., 1996. Phosphorylation dependence of the initiation of productive

transcription of Balbiani ring 2 genes in living cells. *Chromosoma*, 104(6), pp.422–433.
Eick, D. & Geyer, M., 2013. The RNA polymerase II carboxy-terminal domain (CTD) code. *Chemical reviews*, 113(11), pp.8456–8490.
Endoh, M. et al., 2004. Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Molecular and cellular biology*, 24(8), pp.3324–3336.
Eperon, L.P. et al., 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? *CELL*, 54(3), pp.393–401.
Ernst, J. et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), pp.43–49.
Esnault, C. et al., 2008. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Molecular Cell*, 31(3), pp.337–346.
Esnault, C. et al., 2014. Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes & Development*.
FANTOM Consortium, 2014. A promoter-level mammalian expression atlas. *Nature*, 507(7493), pp.462–470.
Fazzio, T.G., Huff, J.T. & Panning, B., 2008. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *CELL*, 134(1), pp.162–174.
Ferreira, M.E. et al., 2005. Mechanism of transcription factor recruitment by acidic activators. *Journal of Biological Chemistry*, 280(23), pp.21779–21784.
Feuer, G. & Molnar, F., 1948. Studies on the composition and polymerization of actin. *Hungarica acta physiologica*, 1(4-5), pp.150–163.
Filtz, T.M., Vogel, W.K. & Leid, M., 2014. Regulation of transcription factor activity by interconnected post- translational modifications. *Trends in Pharmacological Sciences*, 35(2), pp.76–85.
Fischle, W. et al., 2005. Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature*, 438(7071), pp.1116–1122.
Fishburn, J. & Hahn, S., 2012. Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIF. *Molecular and cellular biology*, 32(1), pp.12–25.
FitzGerald, P.C. et al., 2006. Comparative genomics of Drosophila and human core promoters. *Genome biology*, 7(7), p.R53.
Flores, O., Maldonado, E. & Reinberg, D., 1989. Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *The Journal of biological chemistry*, 264(15), pp.8913–8921.
Fowler, T., Sen, R. & Roy, A.L., 2011. Regulation of Primary Response Genes. *Molecular Cell*, 44(3), pp.348–360.
Frangini, A. et al., 2013. The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. *Molecular Cell*, 51(5), pp.647–661.
Fromm, G., Gilchrist, D.A. & Adelman, K., 2013. SnapShot: Transcription regulation: pausing. *CELL*, 153(4), pp.930–930.e1.
Fuchs, S.Y., Tappin, I. & Ronai, Z., 2000. Stability of the ATF2 transcription factor is regulated by phosphorylation and dephosphorylation. *Journal of Biological Chemistry*, 275(17), pp.12560–12564.
Fujinaga, K. et al., 2004. Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Molecular and cellular biology*, 24(2), pp.787–795.
Garber, M. et al., 2012. A high-throughput chromatin immunoprecipitation approach

reveals principles of dynamic gene regulation in mammals. *Molecular Cell*, 47(5), pp.810–822.
Gariglio, P., Bellard, M. & Chambon, P., 1981. Clustering of RNA polymerase B molecules in the 5' moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Research*, 9(11), pp.2589–2598.
Garvie, C.W. & Wolberger, C., 2001. Recognition of specific DNA sequences. *Molecular Cell*, 8(5), pp.937–946.
Gaszner, M. & Felsenfeld, G., 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews Genetics*, 7(9), pp.703–713.
Gazit, K. et al., 2009. TAF4/4b x TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. *The Journal of biological chemistry*, 284(39), pp.26286–26296.
Geertz, M. & Maerkl, S.J., 2010. Experimental strategies for studying transcription factor-DNA binding specificities. *Briefings in functional genomics*, 9(5-6), pp.362–373.
Gerber, A. et al., 2013. Blood-borne circadian signal stimulates daily oscillations in actin dynamics and SRF activity. *CELL*, 152(3), pp.492–503.
Gershenzon, N.I. & Ioshikhes, I.P., 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics (Oxford, England)*, 21(8), pp.1295–1300.
Ghamari, A. et al., 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes & Development*, 27(7), pp.767–777.
Ghazy, M.A. et al., 2009. The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing. *Molecular and cellular biology*, 29(8), pp.2296–2307.
Ghosh, A., Shuman, S. & Lima, C.D., 2011. Structural insights to how mammalian capping enzyme reads the CTD code. *Molecular Cell*, 43(2), pp.299–310.
Giardina, C., Pérez-Riba, M. & Lis, J.T., 1992. Promoter melting and TFIID complexes on Drosophila genes in vivo. *Genes & Development*, 6(11), pp.2190–2200.
Gick, O. et al., 1986. Generation of histone mRNA 3' ends by endonucleolytic cleavage of the pre-mRNA in a snRNP-dependent in vitro reaction. *The EMBO Journal*, 5(6), pp.1319–1326.
Gilbert, W., 1978. Why genes in pieces? *Nature*, 271(5645), p.501.
Gilchrist, D.A. et al., 2008. NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & Development*, 22(14), pp.1921–1933.
Gilchrist, D.A. et al., 2010. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *CELL*, 143(4), pp.540–551.
Gille, H. et al., 1995. ERK phosphorylation potentiates Elk-1-mediated ternary complex formation and transactivation. *The EMBO Journal*, 14(5), pp.951–962.
Gille, H., Sharrocks, A.D. & Shaw, P.E., 1992. Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter. *Nature*, 358(6385), pp.414–417.
Gilmour, D.S. & Lis, J.T., 1986. RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells. *Molecular and cellular biology*, 6(11), pp.3984–3989.
Giniger, E. & Ptashne, M., 1987. Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. *Nature*, 330(6149), pp.670–672.
Giovane, A. et al., 1994. Net, a new ets transcription factor that is activated by Ras. *Genes & Development*, 8(13), pp.1502–1513.

Giridharan, S.S.P. & Caplan, S., 2014. MICAL-family proteins: Complex regulators of the actin cytoskeleton. *Antioxidants & redox signaling*, 20(13), pp.2059–2073.

Glover-Cutter, K. et al., 2009. TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Molecular and cellular biology*, 29(20), pp.5455–5464.

Göke, J. et al., 2013. Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. *Molecular Cell*, 50(6), pp.844–855.

Granger-Schnarr, M. et al., 1992. Transformation and transactivation suppressor activity of the c-Jun leucine zipper fused to a bacterial repressor. *Proceedings of the National Academy of Sciences*, 89(10), pp.4236–4239.

Greenberg, M.E. & Ziff, E.B., 1984. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature*, 311(5985), pp.433–438.

Greenberg, M.E., Greene, L.A. & Ziff, E.B., 1985. Nerve growth factor and epidermal growth factor induce rapid transient changes in proto-oncogene transcription in PC12 cells. *Journal of Biological Chemistry*, 260(26), pp.14101–14110.

Gregory, T.R., 2005. Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics*, 6(9), pp.699–708.

Griner, E.M. & Kazanietz, M.G., 2007. Protein kinase C and other diacylglycerol effectors in cancer. *Nature reviews. Cancer*, 7(4), pp.281–294.

Gromak, N., West, S. & Proudfoot, N.J., 2006. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and cellular biology*, 26(10), pp.3986–3996.

Grünberg, S., Warfield, L. & Hahn, S., 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Structural & Molecular Biology*, 19(8), pp.788–796.

Gu, B., Eick, D. & Bensaude, O., 2013. CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. *Nucleic Acids Research*, 41(3), pp.1591–1603.

Guenther, M.G. et al., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *CELL*, 130(1), pp.77–88.

Guermah, M., Malik, S. & Roeder, R.G., 1998. Involvement of TFIID and USA components in transcriptional activation of the human immunodeficiency virus promoter by NF-kappaB and Sp1. *Molecular and cellular biology*, 18(6), pp.3234–3244.

Guettler, S. et al., 2008. RPEL motifs link the serum response factor cofactor MAL but not myocardin to Rho signaling via actin binding. *Molecular and cellular biology*, 28(2), pp.732–742.

Haberland, M., Montgomery, R.L. & Olson, E.N., 2009. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews Genetics*, 10(1), pp.32–42.

Haberle, V. et al., 2014. Two independent transcription initiation codes overlap on vertebrate core promoters. *Nature*, 507(7492), pp.381–385.

Hah, N. et al., 2011. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *CELL*, 145(4), pp.622–634.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural & Molecular Biology*, 11(5), pp.394–403.

Hall, D.B. & Struhl, K., 2002. The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *Journal of Biological Chemistry*, 277(48), pp.46043–46050.

Hall, J.A. & Georgel, P.T., 2007. CHD proteins: a diverse family with strong ties. *Biochemistry and Cell Biology*, 85(4), pp.463–476.

Hargreaves, D.C. et al., 2009. Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell*, 138(1), pp.129–145.
Hassler, M. & Richmond, T.J., 2001. The B-box dominates SAP-1-SRF interactions in the structure of the ternary complex. *The EMBO Journal*, 20(12), pp.3018–3028.
Hauf, S. et al., 2003. The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *The Journal of cell biology*, 161(2), pp.281–294.
Haupt, Y. et al., 1997. Mdm2 promotes the rapid degradation of p53. *Nature*, 387(6630), pp.296–299.
He, N. et al., 2011. Human polymerase-associated factor complex (PAFc) connects the super elongation complex (SEC) to RNA polymerase II on chromatin. *Proceedings of the National Academy of Sciences*, 108(36), pp.E636–E645.
He, X. et al., 2003. Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. *Genes & Development*, 17(8), pp.1030–1042.
He, Y. et al., 2013. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442), pp.481–486.
Heasman, S.J. & Ridley, A.J., 2008. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nature Reviews Molecular Cell Biology*, 9(9), pp.690–701.
Hermann, S., Berndt, K.D. & Wright, A.P., 2001. How transcriptional activators bind target proteins. *Journal of Biological Chemistry*, 276(43), pp.40127–40132.
Herrera, R.E., Shaw, P.E. & Nordheim, A., 1989. Occupation of the c-fos serum response element in vivo by a multi-protein complex is unaltered by growth factor induction. *Nature*, 340(6228), pp.68–70.
Hill, C.S. & Treisman, R., 1995a. Differential activation of c-fos promoter elements by serum, lysophosphatidic acid, G proteins and polypeptide growth factors. *The EMBO Journal*, 14(20), pp.5037–5047.
Hill, C.S. & Treisman, R., 1995b. Transcriptional regulation by extracellular signals: mechanisms and specificity. *CELL*, 80(2), pp.199–211.
Hill, C.S. et al., 1993. Functional analysis of a growth factor-responsive transcription factor complex. *CELL*, 73(2), pp.395–406.
Hill, C.S., Wynne, J. & Treisman, R., 1994. Serum-regulated transcription by serum response factor (SRF): a novel role for the DNA binding domain. *The EMBO Journal*, 13(22), pp.5421–5432.
Hill, C.S., Wynne, J. & Treisman, R., 1995. The Rho family GTPases RhoA, Rac1, and CDC42Hs regulate transcriptional activation by SRF. *CELL*, 81(7), pp.1159–1170.
Hintermair, C. et al., 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *The EMBO Journal*, 31(12), pp.2784–2797.
Hirose, Y., Tacke, R. & Manley, J.L., 1999. Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes & Development*, 13(10), pp.1234–1239.
Ho, C.K. & Shuman, S., 1999. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Molecular Cell*, 3(3), pp.405–411.
Ho, C.Y. et al., 2013. Lamin A/C and emerin regulate MKL1-SRF activity by modulating actin dynamics. *Nature*, 497(7450), pp.507–511.
Hollenhorst, P.C., McIntosh, L.P. & Graves, B.J., 2011. Genomic and Biochemical Insights into the Specificity of ETS Transcription Factors. *Annual Review of Biochemistry*, 80(1), pp.437–471.
Holliday, R., 2007. Mechanisms for the control of gene activity during development. *Biological reviews of the Cambridge Philosophical Society*, 65(4), pp.431–471.
Howe, L.R. et al., 1992. Activation of the MAP kinase pathway by the protein kinase raf.

CELL, 71(2), pp.335–342.
Hsin, J.-P., Sheth, A. & Manley, J.L., 2011. RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science*, 334(6056), pp.683–686.
Hsin, J.P. & Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development*, 26(19), pp.2119–2137.
Htun, H. et al., 1996. Visualization of glucocorticoid receptor translocation and intranuclear organization in living cells with a green fluorescent protein chimera. *Proceedings of the National Academy of Sciences*, 93(10), pp.4845–4850.
Huang, D.W., Sherman, B.T. & Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1), pp.44–57.
Inman, G.J. & Hill, C.S., 2002. Stoichiometry of active smad-transcription factor complexes on DNA. *Journal of Biological Chemistry*, 277(52), pp.51008–51016.
Jaehning, J.A., 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochimica et biophysica acta*, 1799(5-6), pp.379–388.
Jang, M.K. et al., 2005. The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Molecular Cell*, 19(4), pp.523–534.
Janknecht, R. & Nordheim, A., 1992. Elk-1 protein domains required for direct and SRF-assisted DNA-binding. *Nucleic Acids Research*, 20(13), pp.3317–3324.
Janknecht, R. & Nordheim, A., 1996. MAP kinase-dependent transcriptional coactivation by Elk-1 and its cofactor CBP. *Biochemical and biophysical research communications*, 228(3), pp.831–837.
Janknecht, R., Ernst, W.H., Houthaeve, T., et al., 1993a. C-terminal phosphorylation of the serum-response factor. *European journal of biochemistry / FEBS*, 216(2), pp.469–475.
Janknecht, R., Ernst, W.H., Pingoud, V., et al., 1993b. Activation of ternary complex factor Elk-1 by MAP kinases. *The EMBO Journal*, 12(13), pp.5097–5104.
Jeronimo, C., Bataille, A.R. & Robert, F., 2013. The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chemical reviews*, 113(11), pp.8491–8522.
Jiang, P. et al., 2010. Key roles for MED1 LxxLL motifs in pubertal mammary gland development and luminal-cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(15), pp.6765–6770.
Jiang, Y. et al., 2004. Involvement of transcription termination factor 2 in mitotic repression of transcription elongation. *Molecular Cell*, 14(3), pp.375–385.
Jiao, X. et al., 2013. A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Molecular Cell*, 50(1), pp.104–115.
Jiao, X. et al., 2010. Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature*, 467(7315), pp.608–611.
Jin, C. et al., 2009. H3.3/H2A.Z double variant-containing nucleosomes mark “nucleosome-free regions” of active promoters and other regulatory regions. *Nature genetics*, 41(8), pp.941–945.
Johansen, F.E. & Prywes, R., 1993. Identification of transcriptional activation and inhibitory domains in serum response factor (SRF) by using GAL4-SRF constructs. *Molecular and cellular biology*, 13(8), pp.4640–4647.
Johnsen, S.A., 2012. CDK9 and H2B monoubiquitination: a well-choreographed dance. *PLoS genetics*, 8(8), p.e1002860.
Johnson, G.L. & Lapadat, R., 2002. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science*, 298(5600), pp.1911–1912.

Johnson, K.M. et al., 2002. TFIID and human mediator coactivator complexes assemble cooperatively on promoter DNA. *Genes & Development*, 16(14), pp.1852–1863.

Jolma, A. et al., 2013. DNA-Binding Specificitiesof Human Transcription Factors. *CELL*, 152(1-2), pp.327–339.

Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7), pp.484–492.

Jonker, H.R.A. et al., 2005. Structural properties of the promiscuous VP16 activation domain. *Biochemistry*, 44(3), pp.827–839.

Joseph, R. et al., 2010. Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. *Molecular systems biology*, 6, p.456.

Jurica, M.S. & Moore, M.J., 2003. Pre-mRNA splicing: awash in a sea of proteins. *Molecular Cell*, 12(1), pp.5–14.

Juven-Gershon, T. & Kadonaga, J.T., 2010. Developmental Biology. *Developmental Biology*, 339(2), pp.225–229.

Juven-Gershon, T., Hsu, J.-Y. & Kadonaga, J.T., 2006. Perspectives on the RNA polymerase II core promoter. *Biochemical Society transactions*, 34(Pt 6), pp.1047–1050.

Kagey, M.H. et al., 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314), pp.430–435.

Kaneko, S. et al., 2007. The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3' processing and transcription termination. *Genes & Development*, 21(14), pp.1779–1789.

Kanemaki, M. et al., 1999. TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *The Journal of biological chemistry*, 274(32), pp.22437–22444.

Kao, S.Y. et al., 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature*, 330(6147), pp.489–493.

Kapanidis, A.N. et al., 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802), pp.1144–1147.

Kaplan, T. et al., 2011. Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. *PLoS genetics*, 7(2), p.e1001290.

Karin, M., Liu, Z.G. & Zandi, E., 1997. AP-1 function and regulation. *Current Opinion in Cell Biology*, 9(2), pp.240–246.

Kedinger, C. et al., 1970. Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochemical and biophysical research communications*, 38(1), pp.165–171.

Kim, J., Guermah, M. & Roeder, R.G., 2010. The Human PAF1 Complex Acts in Chromatin Transcription Elongation Both Independently and Cooperatively with SII/TFIIS. *CELL*, 140(4), pp.491–503.

Kim, M. et al., 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *The Journal of biological chemistry*, 284(39), pp.26421–26426.

Kim, M. et al., 2004. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *The EMBO Journal*, 23(2), pp.354–364.

Kim, T.K., Ebright, R.H. & Reinberg, D., 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIH. *Science*, 288(5470), pp.1418–1422.

Kim, Y.J. et al., 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *CELL*, 77(4), pp.599–608.

Kireeva, M.L., Komissarova, N. & Kashlev, M., 2000. Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *Journal of molecular*

biology, 299(2), pp.325–335.
Klemm, J.D., Schreiber, S.L. & Crabtree, G.R., 1998. Dimerization as a regulatory mechanism in signal transduction. *Annual Review of Immunology*, 16, pp.569–592.
Kobayashi, N., Boyer, T.G. & Berk, A.J., 1995. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Molecular and cellular biology*, 15(11), pp.6465–6473.
Kondoh, K. & Nishida, E., 2007. Regulation of MAP kinases by MAP kinase phosphatases. *Biochimica et biophysica acta*, 1773(8), pp.1227–1237.
Kornberg, R.D., 1977. Structure of chromatin. *Annual Review of Biochemistry*, 46, pp.931–954.
Kostrewa, D. et al., 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*, 462(7271), pp.323–330.
Kouhara, H. et al., 1997. A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *CELL*, 89(5), pp.693–702.
Kouzine, F. et al., 2013. Global regulation of promoter melting in naive lymphocytes. *CELL*, 153(5), pp.988–999.
Kriegsheim, von, A. et al., 2009. Cell fate decisions are specified by the dynamic ERK interactome. *Nature cell biology*, 11(12), pp.1458–1464.
Krogan, N.J. et al., 2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Molecular Cell*, 11(3), pp.721–729.
Krumm, A. et al., 1992. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes & Development*, 6(11), pp.2201–2213.
Ku, M. et al., 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics*, 4(10), p.e1000242.
Kummerfeld, S.K. & Teichmann, S.A., 2006. DBD: a transcription factor prediction database. *Nucleic Acids Research*, 34, pp.D74–81.
Kussie, P.H. et al., 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274(5289), pp.948–953.
Kustermans, G. et al., 2005. Perturbation of actin dynamics induces NF-kappaB activation in myelomonocytic cells through an NADPH oxidase-dependent pathway. *The Biochemical journal*, 387(Pt 2), pp.531–540.
Kutach, A.K. & Kadonaga, J.T., 2000. The downstream promoter element DPE appears to be as widely used as the TATA box in Drosophila core promoters. *Molecular and cellular biology*, 20(13), pp.4754–4764.
Kwak, H. & Lis, J.T., 2013. Control of Transcriptional Elongation. *Annual Review of Genetics*, 47(1), pp.483–508.
Kwak, H. et al., 2013. Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science*, 339(6122), pp.950–953.
Laboratories, C., 2013. PT1260-1_072313. pp.1–32.
Lagrange, T. et al., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development*, 12(1), pp.34–44.
Lane, D.P., 1992. Cancer. p53, guardian of the genome. *Nature*, 358(6381), pp.15–16.
Lauberth, S.M. et al., 2013. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *CELL*, 152(5), pp.1021–1036.
Lee, J.-S., Smith, E. & Shilatifard, A., 2010. The Language of Histone Crosstalk. *CELL*, 142(5), pp.682–685.
Lee, K.K. & Workman, J.L., 2007. Histone acetyltransferase complexes: one size doesn't fit all. *Nature Reviews Molecular Cell Biology*, 8(4), pp.284–295.
Lenasi, T., Peterlin, B.M. & Barboric, M., 2011. Cap-binding protein complex links pre-mRNA capping to transcription elongation and alternative splicing through positive

transcription elongation factor b (P-TEFb). *The Journal of biological chemistry*, 286(26), pp.22758–22768.
Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–151.
Li, M.Z. & Elledge, S.J., 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature methods*, 4(3), pp.251–256.
Li, Q.J., 2003. MAP kinase phosphorylation-dependent activation of Elk-1 leads to activation of the co-activator p300. *The EMBO Journal*, 22(2), pp.281–291.
Lifton, R.P. et al., 1978. The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. *Cold Spring Harbor symposia on quantitative biology*, 42 Pt 2, pp.1047–1051.
Lim, C.Y. et al., 2004. The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes & Development*, 18(13), pp.1606–1617.
Lin, C. et al., 2010. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular Cell*, 37(3), pp.429–437.
Ling, Y. et al., 1998. Interaction of transcription factors with serum response factor. Identification of the Elk-1 binding surface. *Journal of Biological Chemistry*, 273(17), pp.10506–10514.
Liu, J. et al., 2006. Intrinsic disorder in transcription factors. *Biochemistry*, 45(22), pp.6873–6888.
Liu, P., Greenleaf, A.L. & Stiller, J.W., 2008. The essential sequence elements required for RNAP II carboxyl-terminal domain function in yeast and their evolutionary conservation. *Molecular Biology and Evolution*, 25(4), pp.719–727.
Lloyd, A., Yancheva, N. & Wasylyk, B., 1991. Transformation suppressor activity of a Jun transcription factor lacking its activation domain. *Nature*, 352(6336), pp.635–638.
Lopez, M. et al., 1994. ERP, a new member of the ets transcription factor/oncoprotein family: cloning, characterization, and differential expression during B-lymphocyte development. *Molecular and cellular biology*, 14(5), pp.3292–3309.
Lorch, Y. et al., 2000. Mediator-nucleosome interaction. *Molecular Cell*, 6(1), pp.197–201.
Lorch, Y., Maier-Davis, B. & Kornberg, R.D., 2006. Chromatin remodeling by nucleosome disassembly in vitro. *Proceedings of the National Academy of Sciences*, 103(9), pp.3090–3093.
Lue, N.F. & Kornberg, R.D., 1987. Accurate initiation at RNA polymerase II promoters in extracts from Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 84(24), pp.8839–8843.
Lundquist, M.R. et al., 2014. Redox modification of nuclear actin by MICAL-2 regulates SRF signaling. *CELL*, 156(3), pp.563–576.
Luo, Z., Lin, C. & Shilatifard, A., 2012. The super elongation complex (SEC) family in transcriptional control. *Nature Reviews Molecular Cell Biology*, 13(9), pp.543–547.
Luscombe, N.M., Laskowski, R.A. & Thornton, J.M., 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Research*, 29(13), pp.2860–2874.
Ma, Z. et al., 2001. Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nature genetics*, 28(3), pp.220–221.
Macdonald, N. et al., 2005. Molecular Basis for the Recognition of Phosphorylated and Phosphoacetylated Histone H3 by 14-3-3. *Molecular Cell*, 20(2), pp.199–211.
Mack, C.P. et al., 2001. Smooth muscle differentiation marker gene expression is regulated by RhoA-mediated actin polymerization. *Journal of Biological Chemistry*, 276(1), pp.341–347.

Macleod, D. et al., 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes & Development*, 8(19), pp.2282–2292.

Madak-Erdogan, Z. et al., 2011. Genomic collaboration of estrogen receptor alpha and extracellular signal-regulated kinase 2 in regulating gene and proliferation programs. *Molecular and cellular biology*, 31(1), pp.226–236.

Mahadevan, L.C., Willis, A.C. & Barratt, M.J., 1991. Rapid histone H3 phosphorylation in response to growth factors, phorbol esters, okadaic acid, and protein synthesis inhibitors. *CELL*, 65(5), pp.775–783.

Maki, C.G., Huibregtse, J.M. & Howley, P.M., 1996. In vivo ubiquitination and proteasome-mediated degradation of p53(1). *Cancer research*, 56(11), pp.2649–2654.

Malek, S. et al., 2001. IkappaBbeta, but not IkappaBalpha, functions as a classical cytoplasmic inhibitor of NF-kappaB dimers by masking both NF-kappaB nuclear localization sequences in resting cells. *Journal of Biological Chemistry*, 276(48), pp.45225–45235.

Malik, S. & Roeder, R.G., 2010. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nature Reviews Genetics*, 11(11), pp.761–772.

Mandal, S.S. et al., 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proceedings of the National Academy of Sciences*, 101(20), pp.7572–7577.

Marais, R., Wynne, J. & Treisman, R., 1993. The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain. *CELL*, 73(2), pp.381–393.

Marais, R.M. et al., 1992. Casein kinase II phosphorylation increases the rate of serum response factor-binding site exchange. *The EMBO Journal*, 11(1), pp.97–105.

Margolis, B. et al., 1999. The function of PTB domain proteins. *Kidney international*, 56(4), pp.1230–1237.

Marshall, C.J., 1988. The ras oncogenes. *Journal of cell science. Supplement*, 10, pp.157–169.

Marzluff, W.F., Wagner, E.J. & Duronio, R.J., 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature Reviews Genetics*, 9(11), pp.843–854.

Matsui, T. et al., 1980. Multiple factors required for accurate initiation of transcription by purified RNA polymerase II. *The Journal of biological chemistry*, 255(24), pp.11992–11996.

Mavrich, T.N. et al., 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Research*, 18(7), pp.1073–1083.

Mayer, A. et al., 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science*, 336(6089), pp.1723–1725.

Mayer, A. et al., 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nature Structural & Molecular Biology*, 17(10), pp.1272–1278.

Mayr, B.M., Canettieri, G. & Montminy, M.R., 2001. Distinct effects of cAMP and mitogenic signals on CREB-binding protein recruitment impart specificity to target gene activation via CREB. *Proceedings of the National Academy of Sciences*, 98(19), pp.10936–10941.

McCracken, S. et al., 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385(6614), pp.357–361.

McDonald, D. et al., 2006. Nucleoplasmic beta-actin exists in a dynamic equilibrium between low-mobility polymeric species and rapidly diffusing populations. *The*

Journal of cell biology, 172(4), pp.541–552.
McGinnis, W. & Krumlauf, R., 1992. Homeobox genes and axial patterning. *CELL*, 68(2), pp.283–302.
McKnight, S.L. et al., 1981. Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. *CELL*, 25(2), pp.385–398.
McManus, C.J. & Graveley, B.R., 2011. RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development*, 21(4), pp.373–379.
Medjkane, S. et al., 2009. Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nature cell biology*, 11(3), pp.257–268.
Meinhart, A. et al., 2005. A structural perspective of CTD function. *Genes & Development*, 19(12), pp.1401–1415.
Mencía, M. et al., 2002. Activator-specific recruitment of TFIID and regulation of ribosomal protein genes in yeast. *Molecular Cell*, 9(4), pp.823–833.
Mendenhall, E.M. et al., 2013. locus-specific editing of histone modifications at endogenous enhancers. *Nature biotechnology*, 31(12), pp.1133–1136.
Mercher, T. et al., 2001. Involvement of a human gene related to the Drosophila spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proceedings of the National Academy of Sciences*, 98(10), pp.5776–5779.
Meredith, G.D. et al., 1996. The C-terminal domain revealed in the structure of RNA polymerase II. *Journal of molecular biology*, 258(3), pp.413–419.
Mermod, N. et al., 1989. The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *CELL*, 58(4), pp.741–753.
Meshorer, E. et al., 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental Cell*, 10(1), pp.105–116.
Meyer, K.D. et al., 2008. Cooperative activity of cdk8 and GCN5L within Mediator directs tandem phosphoacetylation of histone H3. *The EMBO Journal*, 27(10), pp.1447–1457.
Meyer, K.D. et al., 2010. p53 activates transcription by directing structural shifts in Mediator. *Nature Structural & Molecular Biology*, 17(6), pp.753–760.
Meyer, T. & Vinkemeier, U., 2004. Nucleocytoplasmic shuttling of STAT transcription factors. *European journal of biochemistry / FEBS*, 271(23-24), pp.4606–4612.
Michel, M. & Cramer, P., 2013. Transitions for regulating early transcription. *CELL*, 153(5), pp.943–944.
Mikkelsen, T.S. et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), pp.553–560.
Millevoi, S. & Vagner, S., 2010. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Research*, 38(9), pp.2757–2774.
Min, I.M. et al., 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & Development*, 25(7), pp.742–754.
Miralles, F. et al., 2003. Actin dynamics control SRF activity by regulation of its coactivator MAL. *CELL*, 113(3), pp.329–342.
Missra, A. & Gilmour, D.S., 2010. Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), pp.11301–11306.
Mitchell, P.J. & Tjian, R., 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, 245(4916), pp.371–378.
Mittler, G. et al., 2003. A novel docking site on Mediator is critical for activation by VP16 in mammalian cells. *The EMBO Journal*, 22(24), pp.6494–6504.
Mizutani, A. & Tanaka, M., 2003. Regions of GAL4 critical for binding to a promoter in

vivo revealed by a visual DNA-binding analysis. *The EMBO Journal*, 22(9), pp.2178–2187.
Mohan, M., Herz, H.-M. & Shilatifard, A., 2012. SnapShot: Histone Lysine Methylase Complexes. *CELL*, 149(2), pp.498–498.e1.
Mohun, T., Garrett, N. & Treisman, R., 1987. Xenopus cytoskeletal actin and human c-fos gene promoters share a conserved protein-binding site. *The EMBO Journal*, 6(3), pp.667–673.
Moon, R.T. & Kimelman, D., 1998. From cortical rotation to organizer gene expression: toward a molecular explanation of axis specification in Xenopus. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 20(7), pp.536–545.
Moore, C.L. & Sharp, P.A., 1985. Accurate cleavage and polyadenylation of exogenous RNA substrate. *CELL*, 41(3), pp.845–855.
Moore, M.J. & Proudfoot, N.J., 2009. Pre-mRNA processing reaches back to transcription and ahead to translation. *CELL*, 136(4), pp.688–700.
Morey, L. & Helin, K., 2010. Polycomb group protein-mediated repression of transcription. *Trends in Biochemical Sciences*, 35(6), pp.323–332.
Morita, T. & Hayashi, K., 2014. Arp5 is a key regulator of myocardin in smooth muscle cells. *The Journal of cell biology*, 204(5), pp.683–696.
Mosammaparast, N. & Shi, Yang, 2010. Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annual Review of Biochemistry*, 79, pp.155–179.
Mouilleron, S. et al., 2008. Molecular basis for G-actin binding to RPEL motifs from the serum response factor coactivator MAL. *The EMBO Journal*, 27(23), pp.3198–3208.
Mouilleron, S. et al., 2011. Structure of a pentavalent G-actin*MRTF-A complex reveals how G-actin controls nucleocytoplasmic shuttling of a transcriptional coactivator. *Science Signaling*, 4(177), p.ra40.
Mueller, C.G. & Nordheim, A., 1991. A protein domain conserved between yeast MCM1 and human SRF directs ternary complex formation. *The EMBO Journal*, 10(13), pp.4219–4229.
Muñoz, M.J. et al., 2009. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *CELL*, 137(4), pp.708–720.
Murai, K. & Treisman, R., 2002. Interaction of serum response factor (SRF) with the Elk-1 B box inhibits RhoA-actin signaling to SRF and potentiates transcriptional activation by Elk-1. *Molecular and cellular biology*, 22(20), pp.7083–7092.
Muse, G.W. et al., 2007. RNA polymerase is poised for activation across the genome. *Nature genetics*, 39(12), pp.1507–1511.
Naar, A.M. et al., 2002. Human CRSP interacts with RNA polymerase II CTD and adopts a specific CTD-bound conformation. *Genes & Development*, 16(11), pp.1339–1344.
Narita, T. et al., 2003. Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex. *Molecular and cellular biology*, 23(6), pp.1863–1873.
Nechaev, S. et al., 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science*, 327(5963), pp.335–338.
Nedea, E. et al., 2003. Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. *The Journal of biological chemistry*, 278(35), pp.33000–33010.
Nemet, J. et al., 2014. The two faces of Cdk8, a positive/negative regulator of transcription. *Biochimie*, 97(c), pp.22–27.
Nguyen, V.T. et al., 2001. 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature*, 414(6861), pp.322–325.

Nojima, T. et al., 2013. Definition of RNA polymerase II CoTC terminator elements in the human genome. *Cell reports*, 3(4), pp.1080–1092.

Nordick, K. et al., 2008. Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II. *Eukaryotic cell*, 7(7), pp.1158–1167.

Norman, C. et al., 1988. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. *CELL*, 55(6), pp.989–1003.

Nurrish, S.J. & Treisman, R., 1995. DNA binding specificity determinants in MADS-box transcription factors. *Molecular and cellular biology*, 15(8), pp.4076–4085.

Odom, D.T. et al., 2007. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature genetics*, 39(6), pp.730–732.

Ohler, U. et al., 2002. Computational analysis of core promoters in the Drosophila genome. *Genome biology*, 3(12), p.research0087.

Ohtsuki, K. et al., 2010. Genome-wide localization analysis of a complete set of Tafs reveals a specific effect of the taf1 mutation on Taf2 occupancy and provides indirect evidence for different TFIID conformations at different promoters. *Nucleic Acids Research*, 38(6), pp.1805–1820.

Orphanides, G. et al., 1998. FACT, a factor that facilitates transcript elongation through nucleosomes. *CELL*, 92(1), pp.105–116.

Ozsolak, F. et al., 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *CELL*, 143(6), pp.1018–1029.

Pal, M., Ponticelli, A.S. & Luse, D.S., 2005. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Molecular Cell*, 19(1), pp.101–110.

Pan, G. & Greenblatt, J., 1994. Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *The Journal of biological chemistry*, 269(48), pp.30101–30104.

Park, S.W. et al., 2005. Thyroid hormone-induced juxtaposition of regulatory elements/factors and chromatin remodeling of Crabp1 dependent on MED1/TRAP220. *Molecular Cell*, 19(5), pp.643–653.

Passmore, S. et al., 1988. Saccharomyces cerevisiae protein involved in plasmid maintenance is necessary for mating of MAT alpha cells. *Journal of molecular biology*, 204(3), pp.593–606.

Patel, D.J. & Wang, Zhanxin, 2013. Readout of Epigenetic Modifications. *Annual Review of Biochemistry*, 82(1), pp.81–118.

Pawson, T., 1995. Protein modules and signalling networks. *Nature*, 373(6515), pp.573–580.

Pawson, T., 2004. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *CELL*, 116(2), pp.191–203.

Pawłowski, R. et al., 2010. An actin-regulated importin α / β -dependent extended bipartite NLS directs nuclear import of MRTF-A. *The EMBO Journal*, 29(20), pp.3448–3458.

Pearson, J.C., Lemons, D. & McGinnis, W., 2005. Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics*, 6(12), pp.893–904.

Pederson, T. & Aebi, U., 2002. Actin in the nucleus: what form and what for? *Journal of structural biology*, 140(1-3), pp.3–9.

Pei, Y., Schwer, B. & Shuman, S., 2003. Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *Journal of Biological Chemistry*, 278(9), pp.7180–7188.

Pelka, P. et al., 2008. Intrinsic structural disorder in adenovirus E1A: a viral molecular

hub linking multiple diverse processes. *Journal of virology*, 82(15), pp.7252–7263.
Pellegrini, L., Tan, S. & Richmond, T.J., 1995. Structure of serum response factor core bound to DNA. *Nature*, 376(6540), pp.490–498.
Peng, J., Marshall, N.F. & Price, D.H., 1998. Identification of a cyclin subunit required for the function of Drosophila P-TEFb. *The Journal of biological chemistry*, 273(22), pp.13855–13860.
Percipalle, P., 2013. Co-transcriptional nuclear actin dynamics. *Nucleus (Austin, Tex.)*, 4(1), pp.43–52.
Perlmann, T. & Wrange, O., 1988. Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. *The EMBO Journal*, 7(10), pp.3073–3079.
Petrykowska, H.M., Vockley, C.M. & Elnitski, L., 2008. Detection and characterization of silencers and enhancer-blockers in the greater CFTR locus. *Genome Research*, 18(8), pp.1238–1246.
Pierreux, C.E., Nicolás, F.J. & Hill, C.S., 2000. Transforming growth factor beta-independent shuttling of Smad4 between the cytoplasm and nucleus. *Molecular and cellular biology*, 20(23), pp.9041–9054.
Pipes, G.C.T., Creemers, E.E. & Olson, E.N., 2006. The myocardin family of transcriptional coactivators: versatile regulators of cell growth, migration, and myogenesis. *Genes & Development*, 20(12), pp.1545–1556.
Pires-daSilva, A. & Sommer, R.J., 2003. The evolution of signalling pathways in animal development. *Nature Reviews Genetics*, 4(1), pp.39–49.
Plant, K.E. et al., 2005. Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene. *Molecular and cellular biology*, 25(8), pp.3276–3285.
Plotnikov, A. et al., 2011. The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et biophysica acta*, 1813(9), pp.1619–1633.
Pollard, T.D., 2007. Regulation of actin filament assembly by Arp2/3 complex and formins. *Annual review of biophysics and biomolecular structure*, 36, pp.451–477.
Pollard, T.D. & Borisy, G.G., 2003. Cellular motility driven by assembly and disassembly of actin filaments. *CELL*, 112(4), pp.453–465.
Pollard, T.D. & Cooper, J.A., 2009. Actin, a Central Player in Cell Shape and Movement. *Science*, 326(5957), pp.1208–1212.
Pollock, R. & Treisman, R., 1990. A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Research*, 18(21), pp.6197–6204.
Pollock, R. & Treisman, R., 1991. Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes & Development*, 5(12A), pp.2327–2341.
Poorey, K. et al., 2013. Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science*, 342(6156), pp.369–372.
Posern, G., Sotiropoulos, A. & Treisman, R., 2002. Mutant actins demonstrate a role for unpolymerized actin in control of transcription by serum response factor. *Molecular biology of the cell*, 13(12), pp.4167–4178.
Preker, P. et al., 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(5909), pp.1851–1854.
Price, D.H., 2000. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Molecular and cellular biology*, 20(8), pp.2629–2634.
Price, D.H., Sluder, A.E. & Greenleaf, A.L., 1989. Dynamic interaction between a Drosophila transcription factor and RNA polymerase II. *Molecular and cellular biology*, 9(4), pp.1465–1475.
Price, M.A., Rogers, A.E. & Treisman, R., 1995. Comparative analysis of the ternary complex factors Elk-1, SAP-1a and SAP-2 (ERP/NET). *The EMBO Journal*, 14(11), pp.2589–2601.

Primig, M., Winkler, H. & Ammerer, G., 1991. The DNA binding and oligomerization domain of MCM1 is sufficient for its interaction with other regulatory proteins. *The EMBO Journal*, 10(13), pp.4209–4218.

Proudfoot, N.J., 2011. Ending the message: poly(A) signals then and now. *Genes & Development*, 25(17), pp.1770–1782.

Proudfoot, N.J., 1989. How RNA polymerase II terminates transcription in higher eukaryotes. *Trends in Biochemical Sciences*, 14(3), pp.105–110.

Proudfoot, N.J. & Brownlee, G.G., 1976. 3' non-coding region sequences in eukaryotic messenger RNA. *Nature*, 263(5574), pp.211–214.

Pruitt, K.D. et al., 2013. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research*, 42(D1), pp.D756–D763.

Prywes, R. & Roeder, R.G., 1987. Purification of the c-fos enhancer-binding protein. *Molecular and cellular biology*, 7(10), pp.3482–3489.

Ptashne, M. & Gann, A., 1997. Transcriptional activation by recruitment. *Nature*.

Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D., 2011. RAS oncogenes: weaving a tumorigenic web. *Nature reviews. Cancer*, 11(11), pp.761–774.

Rahl, P.B. et al., 2010. c-Myc Regulates Transcriptional Pause Release. *CELL*, 141(3), pp.432–445.

Ram, O. et al., 2011. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *CELL*, 147(7), pp.1628–1639.

Rao, V.N. et al., 1989. elk, tissue-specific ets-related genes on chromosomes X and 14 near translocation breakpoints. *Science*, 244(4900), pp.66–70.

Ravasi, T. et al., 2010. An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *CELL*, 140(5), pp.744–752.

Raveh-Sadka, T., Levo, M. & Segal, E., 2009. Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Research*, 19(8), pp.1480–1496.

Ray, L.B. & Sturgill, T.W., 1988. Insulin-stimulated microtubule-associated protein kinase is phosphorylated on tyrosine and threonine in vivo. *Proceedings of the National Academy of Sciences*, 85(11), pp.3753–3757.

Reeve, J.N., 2003. Archaeal chromatin and transcription. *Molecular microbiology*, 48(3), pp.587–598.

Regier, J.L., Shen, F. & Triezenberg, S.J., 1993. Pattern of aromatic and hydrophobic amino acids critical for one of two subdomains of the VP16 transcriptional activator. *Proceedings of the National Academy of Sciences*, 90(3), pp.883–887.

Remedios, dos, C.G. et al., 2003. Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiological reviews*, 83(2), pp.433–473.

Reppas, N.B. et al., 2006. The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Molecular Cell*, 24(5), pp.747–757.

Revyakin, A. et al., 2006. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, 314(5802), pp.1139–1143.

Rhee, H.S. & Pugh, B.F., 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483(7389), pp.295–301.

Rice, A.P., 2010. The HIV-1 Tat team gets bigger. *Cell host & microbe*, 7(3), pp.179–181.

Riising, E.M. et al., 2014. Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular Cell*, pp.1–48.

Roberts, G.C. et al., 1998. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Research*, 26(24), pp.5568–5572.

Robinson, P.J.J. et al., 2012. Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), pp.17931–17935.

Roeder, R.G. & Rutter, W.J., 1969. Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature*, 224(5216), pp.234–237.

Roeder, R.G. & Rutter, W.J., 1970. Specific nucleolar and nucleoplasmic RNA polymerases. *Proceedings of the National Academy of Sciences*, 65(3), pp.675–682.

Rogozin, I.B. et al., 2012. Origin and evolution of spliceosomal introns. *Biology direct*, 7, p.11.

Rohs, R. et al., 2010. Origins of Specificity in Protein-DNA Recognition. *Annual Review of Biochemistry*, 79(1), pp.233–269.

Roskoski, R., 2012a. ERK1/2 MAP kinases: structure, function, and regulation. *Pharmacological research : the official journal of the Italian Pharmacological Society*, 66(2), pp.105–143.

Roskoski, R., 2012b. MEK1/2 dual-specificity protein kinases: structure and regulation. *Biochemical and biophysical research communications*, 417(1), pp.5–10.

Rossetto, D., Avvakumov, N. & Côté, J., 2012. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics : official journal of the DNA Methylation Society*, 7(10), pp.1098–1108.

Rougvie, A.E. & Lis, J.T., 1990. Postinitiation transcriptional control in Drosophila melanogaster. *Molecular and cellular biology*, 10(11), pp.6041–6045.

Rougvie, A.E. & Lis, J.T., 1988. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *CELL*, 54(6), pp.795–804.

Ruhl, D.D. et al., 2006. Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. *Biochemistry*, 45(17), pp.5671–5677.

Ruthenburg, A.J., Allis, C.D. & Wysocka, J., 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular Cell*, 25(1), pp.15–30.

Sadowski, I. et al., 1988. GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, 335(6190), pp.563–564.

Saez, E. et al., 1995. c-fos is required for malignant progression of skin tumors. *CELL*, 82(5), pp.721–732.

Sainsbury, S., Niesser, J. & Cramer, P., 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature*, 493(7432), pp.437–440.

Salinas, S. et al., 2004. SUMOylation regulates nucleo-cytoplasmic shuttling of Elk-1. *The Journal of cell biology*, 165(6), pp.767–773.

Sato, S. et al., 2004. A set of consensus mammalian mediator subunits identified by multidimensional protein identification technology. *Molecular Cell*, 14(5), pp.685–691.

Sawicka, A. & Seiser, C., 2012. Histone H3 phosphorylation. *Biochimie*, 94(11), pp.2193–2201.

Schaefer, U., Schmeier, S. & Bajic, V.B., 2011. TcoF-DB: dragon database for human transcription co-factors and transcription factor interacting proteins. *Nucleic Acids Research*, 39, pp.D106–10.

Schaufele, F. et al., 1986. Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. *Nature*, 323(6091), pp.777–781.

Scheer, U. et al., 1984. Microinjection of actin-binding proteins and actin antibodies demonstrates involvement of nuclear actin in transcription of lampbrush chromosomes. *CELL*, 39(1), pp.111–122.

Schlessinger, J., 2003. Signal transduction. Autoinhibition control. *Science*, 300(5620), pp.750–752.

Schlessinger, J. & Bar-Sagi, D., 1994. Activation of Ras and other signaling pathways

by receptor tyrosine kinases. *Cold Spring Harbor symposia on quantitative biology*, 59, pp.173–179.
Schmidt, A. & Hall, A., 2002. Guanine nucleotide exchange factors for Rho GTPases: turning on the switch. *Genes & Development*, 16(13), pp.1587–1609.
Schones, D.E. et al., 2008. Dynamic regulation of nucleosome positioning in the human genome. *CELL*, 132(5), pp.887–898.
Schratt, G. et al., 2001. Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. *Molecular and cellular biology*, 21(8), pp.2933–2943.
Schröter, H. et al., 1990. Synergism in ternary complex formation between the dimeric glycoprotein p67SRF, polypeptide p62TCF and the c-fos serum response element. *The EMBO Journal*, 9(4), pp.1123–1130.
Schümperli, D., 1988. Multilevel regulation of replication-dependent histone genes. *Trends in genetics : TIG*, 4(7), pp.187–191.
Schwer, B., Sanchez, A.M. & Shuman, S., 2012. Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), pp.18024–18029.
Segal, E. et al., 2006. A genomic code for nucleosome positioning. *Nature*, 442(7104), pp.772–778.
Seizl, M. et al., 2011. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic Acids Research*, 39(14), pp.6291–6304.
Sekinger, E.A., Moqtaderi, Z. & Struhl, K., 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Molecular Cell*, 18(6), pp.735–748.
Selvaraj, A. & Prywes, R., 2003. Megakaryoblastic leukemia-1/2, a transcriptional co-activator of serum response factor, is required for skeletal myogenic differentiation. *The Journal of biological chemistry*, 278(43), pp.41977–41987.
Sharrocks, A.D., 2002. Complexities in ETS-domain transcription factor function and regulation: lessons from the TCF (ternary complex factor) subfamily. The Colworth Medal Lecture. *Biochemical Society transactions*, 30(2), pp.1–9.
Sharrocks, A.D., 2001. The ETS-domain transcription factor family. *Nature Reviews Molecular Cell Biology*, 2(11), pp.827–837.
Sharrocks, A.D., Hesler, von, F. & Shaw, P.E., 1993. The identification of elements determining the different DNA binding specificities of the MADS box proteins p67SRF and RSRFC4. *Nucleic Acids Research*, 21(2), pp.215–221.
Shaulian, E. & Karin, M., 2001. AP-1 in cell proliferation and survival. *Oncogene*, 20(19), pp.2390–2400.
Shaw, P.E., 1992. Ternary complex formation over the c-fos serum response element: p62TCF exhibits dual component specificity with contacts to DNA and an extended structure in the DNA-binding domain of p67SRF. *The EMBO Journal*, 11(8), pp.3011–3019.
Shaw, P.E., Schröter, H. & Nordheim, A., 1989. The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. *CELL*, 56(4), pp.563–572.
Sheldon, K.E., Mauger, D.M. & Arndt, K.M., 2005. A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation. *Molecular Cell*, 20(2), pp.225–236.
Shen, Y. et al., 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409), pp.116–120.
Shi, X et al., 1996. Paf1p, an RNA polymerase II-associated factor in Saccharomyces cerevisiae, may have both positive and negative roles in transcription. *Molecular and cellular biology*, 16(2), pp.669–676.

Shi, Yigong & Massagué, J., 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *CELL*, 113(6), pp.685–700.

Shi, Yujiang et al., 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *CELL*, 119(7), pp.941–953.

Shiekhattar, R. et al., 1995. Cdk-activating kinase complex is a component of human transcription factor TFIIH. *Nature*, 374(6519), pp.283–287.

Shilatifard, A., 2006. Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual Review of Biochemistry*, 75, pp.243–269.

Shimojo, H., Ohtsuka, T. & Kageyama, R., 2008. Oscillations in notch signaling regulate maintenance of neural progenitors. *Neuron*, 58(1), pp.52–64.

Shlyueva, D., Stampfel, G. & Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews Genetics*, 15(4), pp.272–286.

Shore, P. & Sharrocks, A.D., 1995. The MADS-box family of transcription factors. *European journal of biochemistry / FEBS*, 229(1), pp.1–13.

Shore, P. & Sharrocks, A.D., 1994. The transcription factors Elk-1 and serum response factor interact by direct protein-protein contacts mediated by a short region of Elk-1. *Molecular and cellular biology*, 14(5), pp.3283–3291.

Sims, R.J. & Reinberg, D., 2008. Is there a code embedded in proteins that is based on post-translational modifications? *Nature Reviews Molecular Cell Biology*, 9(10), pp.815–820.

Singh, N. & Han, M., 1995. sur-2, a novel gene, functions late in the let-60 ras-mediated signaling pathway during Caenorhabditis elegans vulval induction. *Genes & Development*, 9(18), pp.2251–2265.

Skourti-Stathaki, K., Proudfoot, N.J. & Gromak, N., 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular Cell*, 42(6), pp.794–805.

Small, S., Blair, A. & Levine, M., 1992. Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO Journal*, 11(11), pp.4047–4057.

Smith, E. & Shilatifard, A., 2010. The Chromatin Signaling Pathway: Diverse Mechanisms of Recruitment of Histone-Modifying Enzymes and Varied Biological Outcomes. *Molecular Cell*, 40(5), pp.689–701.

Smith, Z.D. & Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, 14(3), pp.204–220.

Song, A. et al., 1998. Phosphorylation of nuclear MyoD is required for its rapid degradation. *Molecular and cellular biology*, 18(9), pp.4994–4999.

Sotiropoulos, A. et al., 1999. Signal-regulated activation of serum response factor is mediated by changes in actin dynamics. *CELL*, 98(2), pp.159–169.

Soufi, A., Donahue, G. & Zaret, K.S., 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *CELL*, 151(5), pp.994–1004.

Soutourina, J. et al., 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science*, 331(6023), pp.1451–1454.

Spitz, F. & Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*, 13(9), pp.613–626.

Sprang, S.R. et al., 1991. Structural basis for the activation of glycogen phosphorylase b by adenosine monophosphate. *Science*, 254(5036), pp.1367–1371.

Spudich, J.A. & Watt, S., 1971. The regulation of rabbit skeletal muscle contraction. I. Biochemical studies of the interaction of the tropomyosin-troponin complex with actin and the proteolytic fragments of myosin. *The Journal of biological chemistry*, 246(15), pp.4866–4871.

St Amour, C.V. et al., 2012. Separate domains of fission yeast Cdk9 (P-TEFb) are

required for capping enzyme recruitment and primed (Ser7-phosphorylated) Rpb1 carboxyl-terminal domain substrate recognition. *Molecular and cellular biology*, 32(13), pp.2372–2383.
Stent, G.S., 1965. 1965 Nobel laureates in medicine or physiology. *Science*, 150(3695), pp.462–464.
Sterner, D.E. & Berger, S.L., 2000. Acetylation of histones and transcription-related factors. *Microbiology and Molecular Biology Reviews*, 64(2), pp.435–459.
Stevens, J.L. et al., 2002. Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science*, 296(5568), pp.755–758.
Sullivan, A.L. et al., 2011. Serum response factor utilizes distinct promoter- and enhancer-based mechanisms to regulate cytoskeletal gene expression in macrophages. *Molecular and cellular biology*, 31(4), pp.861–875.
Sun, M. et al., 2010. A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *The Journal of biological chemistry*, 285(53), pp.41597–41603.
Sun, X.J. et al., 1993. Pleiotropic insulin signals are engaged by multisite phosphorylation of IRS-1. *Molecular and cellular biology*, 13(12), pp.7418–7428.
Takagaki, Y., Ryner, L.C. & Manley, J.L., 1988. Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation. *CELL*, 52(5), pp.731–742.
Takagi, Y. & Kornberg, R.D., 2006. Mediator as a general transcription factor. *The Journal of biological chemistry*, 281(1), pp.80–89.
Takahashi, H. et al., 2011. Human mediator subunit MED26 functions as a docking site for transcription elongation factors. *CELL*, 146(1), pp.92–104.
Tam, W.F. et al., 2000. Cytoplasmic sequestration of rel proteins by IkappaBalpha requires CRM1-dependent nuclear export. *Molecular and cellular biology*, 20(6), pp.2269–2284.
Tan, M. et al., 2011. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *CELL*, 146(6), pp.1016–1028.
Tatarakis, A. et al., 2008. Dominant and redundant functions of TFIID involved in the regulation of hepatic genes. *Molecular Cell*, 31(4), pp.531–543.
Taylor, G.C.A. et al., 2013. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Research*, 23(12), pp.2053–2065.
Taylor, M. et al., 1989. Muscle-specific (CArG) and serum-responsive (SRE) promoter elements are functionally interchangeable in Xenopus embryos and mouse fibroblasts. *Development (Cambridge, England)*, 106(1), pp.67–78.
Tcherkezian, J. & Lamarche-Vane, N., 2007. Current knowledge of the large RhoGAP family of proteins. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 99(2), pp.67–86.
Thomas, M.C. & Chiang, C.-M., 2006. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, 41(3), pp.105–178.
Thompson, C.M. & Young, R.A., 1995. General requirement for RNA polymerase II holoenzymes in vivo. *Proceedings of the National Academy of Sciences*, 92(10), pp.4587–4590.
Thompson, C.M. et al., 1993. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *CELL*, 73(7), pp.1361–1375.
Thurman, R.E. et al., 2012. The accessible chromatin landscape of the human genome. *Nature*, 488(7414), pp.75–82.
Tian, Z. et al., 2012. Enhanced top-down characterization of histone post-translational modifications. *Genome biology*, 13(10), p.R86.
Tippmann, S.C. et al., 2012. Chromatin measurements reveal contributions of

synthesis and decay to steady-state mRNA levels. *Molecular systems biology*, 8, p.593.

Tirode, F. et al., 1999. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Molecular Cell*, 3(1), pp.87–95.

Tiwari, V.K. et al., 2012. A chromatin-modifying function of JNK during stem cell differentiation. *Nature genetics*, 44(1), pp.94–100.

Treisman, R., 1995a. DNA-binding proteins. Inside the MADS box. *Nature*, 376(6540), pp.468–469.

Treisman, R., 1986. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *CELL*, 46(4), pp.567–574.

Treisman, R., 1995b. Journey to the surface of the cell: Fos regulation and the SRE. *The EMBO Journal*, 14(20), pp.4905–4913.

Treisman, R., 1994. Ternary complex factors: growth factor regulated transcriptional activators. *Current Opinion in Genetics & Development*, 4(1), pp.96–101.

Treisman, R., 1985. Transient accumulation of c-fos RNA following serum stimulation requires a conserved 5“ element and c-fos 3” sequences. *CELL*, 42(3), pp.889–902.

Treisman, R., Marais, R. & Wynne, J., 1992. Spatial flexibility in ternary complexes between SRF and its accessory proteins. *The EMBO Journal*, 11(12), pp.4631–4640.

Triezenberg, S.J., Kingsbury, R.C. & McKnight, S.L., 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes & Development*, 2(6), pp.718–729.

Tsai, K.-L. et al., 2013. A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature Structural & Molecular Biology*, 20(5), pp.611–619.

Tsai, K.-L. et al., 2014. Subunit Architecture and Functional Modular Rearrangementsof the Transcriptional Mediator Complex. *CELL*, 157(6), pp.1430–1444.

Uesugi, M. & Verdine, G.L., 1999. The alpha-helical FXXPhiPhi motif in p53: TAF interaction and discrimination by MDM2. *Proceedings of the National Academy of Sciences*, 96(26), pp.14801–14806.

Uesugi, M. et al., 1997. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science*, 277(5330), pp.1310–1313.

Ullrich, A. & Schlessinger, J., 1990. Signal transduction by receptors with tyrosine kinase activity. *CELL*, 61(2), pp.203–212.

Valekunja, U.K. et al., 2013. Histone methyltransferase MLL3 contributes to genome-scale circadian transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4), pp.1554–1559.

Valouev, A. et al., 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature methods*, 5(9), pp.829–834.

van Ingen, H. et al., 2008. Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure (London, England : 1993)*, 16(8), pp.1245–1256.

van Nimwegen, E., 2003. Scaling laws in the functional content of genomes. *Trends in genetics : TIG*, 19(9), pp.479–484.

Vandel, L. & Kouzarides, T., 1999. Residues phosphorylated by TFIIH are required for E2F-1 degradation during S-phase. *The EMBO Journal*, 18(15), pp.4280–4291.

Vartiainen, M.K. et al., 2007. Nuclear Actin Regulates Dynamic Subcellular Localization and Activity of the SRF Cofactor MAL. *Science*, 316(5832), pp.1749–1752.

Vermeulen, M. et al., 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *CELL*, 142(6), pp.967–980.

Vicent, G.P. et al., 2011. Four enzymes cooperate to displace histone H1 during the

first minute of hormonal gene activation. *Genes & Development*, 25(8), pp.845–862.
Vicent, G.P. et al., 2006. Induction of Progesterone Target Genes Requires Activation of Erk and Msk Kinases and Phosphorylation of Histone H3. *Molecular Cell*, 24(3), pp.367–381.
Viladevall, L. et al., 2009. TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. *Molecular Cell*, 33(6), pp.738–751.
Villar, D., Flicek, P. & Odom, D.T., 2014. Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nature Reviews Genetics*, 15(4), pp.221–233.
Voigt, P., Tee, W.W. & Reinberg, D., 2013. A double take on bivalent promoters. *Genes & Development*, 27(12), pp.1318–1338.
Vojnic, E. et al., 2011. Structure and VP16 binding of the Mediator Med25 activator interaction domain. *Nature Structural & Molecular Biology*, 18(4), pp.404–409.
Voss, T.C. et al., 2011. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *CELL*, 146(4), pp.544–554.
Wada, T. et al., 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development*, 12(3), pp.343–356.
Waddington, C., 1959. Canalization of development and genetic assimilation of acquired characters. *Nature*, 183(4676), pp.1654–1655.
Walker, S., Greaves, R. & O'Hare, P., 1993. Transcriptional activation by the acidic domain of Vmw65 requires the integrity of the domain and involves additional determinants distinct from those necessary for TFIIB binding. *Molecular and cellular biology*, 13(9), pp.5233–5244.
Walter, W. et al., 2008. 14-3-3 interaction with histone H3 involves a dual modification pattern of phosphoacetylation. *Molecular and cellular biology*, 28(8), pp.2840–2849.
Wang, D. et al., 2001a. Activation of cardiac gene expression by myocardin, a transcriptional cofactor for serum response factor. *CELL*, 105(7), pp.851–862.
Wang, D.Z. et al., 2002. Potentiation of serum response factor activity by a family of myocardin-related transcription factors. *Proceedings of the National Academy of Sciences*, 99(23), pp.14855–14860.
Wang, G. et al., 2001b. Characterization of mediator complexes from HeLa cell nuclear extract. *Molecular and cellular biology*, 21(14), pp.4604–4613.
Wang, G. et al., 2005. Mediator Requirement for Both Recruitment and Postrecruitment Steps in Transcription Initiation. *Molecular Cell*, 17(5), pp.683–694.
Wang, Zhigao et al., 2004. Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression. *Nature*, 428(6979), pp.185–189.
Wang, Zhigao et al., 2003. Myocardin is a master regulator of smooth muscle gene expression. *Proceedings of the National Academy of Sciences*, 100(12), pp.7129–7134.
Wei, G.-H. et al., 2010. Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *The EMBO Journal*, 29(13), pp.2147–2160.
Weil, P.A. et al., 1979. Selective and accurate initiation of transcription at the Ad2 major late promotor in a soluble system dependent on purified RNA polymerase II and DNA. *CELL*, 18(2), pp.469–484.
Weiss, B.S. & Gladstone, L., 1959. A mammalian system for the incorporation of cytidinetriphosphate into ribonucleic acid. *Journal of the American Chemical Society*.
Wen, Y. & Shatkin, A.J., 1999. Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes & Development*, 13(14), pp.1774–1779.
West, S., Gromak, N. & Proudfoot, N.J., 2004. Human 5“ --> 3” exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*,

432(7016), pp.522–525.
West, S., Proudfoot, N.J. & Dye, M.J., 2008. Molecular dissection of mammalian RNA polymerase II transcriptional termination. *Molecular Cell*, 29(5), pp.600–610.
Whitelaw, E. & Proudfoot, N., 1986. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO Journal*, 5(11), pp.2915–2922.
Widmann, C. et al., 1999. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiological reviews*, 79(1), pp.143–180.
Wilson, M.D. et al., 2008. Species-specific transcription in mice carrying human chromosome 21. *Science*, 322(5900), pp.434–438.
Wingender, E., Schoeps, T. & Dönitz, J., 2013. TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Research*, 41, pp.D165–70.
Winter, S. et al., 2008. 14-3-3 proteins recognize a histone code at histone H3 and are required for transcriptional activation. *The EMBO Journal*, 27(1), pp.88–99.
Wittschieben, B.O. et al., 1999. A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Molecular Cell*, 4(1), pp.123–128.
Wong, K.H., Jin, Y. & Struhl, K., 2014. TFIIH Phosphorylation of the Pol II CTD Stimulates Mediator Dissociation from the Preinitiation Complex and Promoter Escape. *Molecular Cell*.
Wood, A. et al., 2003. The Paf1 complex is essential for histone monoubiquitination by the Rad6-Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *The Journal of biological chemistry*, 278(37), pp.34739–34742.
Wu, Y. et al., 2010. Structure of the MADS-box/MEF2 domain of MEF2A bound to DNA and its implication for myocardin recruitment. *Journal of molecular biology*, 397(2), pp.520–533.
Wynne, J. & Treisman, R., 1992. SRF and MCM1 have related but distinct DNA binding specificities. *Nucleic Acids Research*, 20(13), pp.3297–3303.
Xiang, K. et al., 2010. Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex. *Nature*, 467(7316), pp.729–733.
Xiao, H. et al., 1994. Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Molecular and cellular biology*, 14(10), pp.7013–7024.
Xin, H. et al., 2009. yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement. *Molecular Cell*, 35(3), pp.365–376.
Xu, L. et al., 2002. Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. *Molecular Cell*, 10(2), pp.271–282.
Yamada, T. et al., 2006. P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Molecular Cell*, 21(2), pp.227–237.
Yamaguchi, Y. et al., 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *CELL*, 97(1), pp.41–51.
Yang, F. et al., 2006. An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*, 442(7103), pp.700–704.
Yang, S.-H. et al., 2003. Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. *Molecular Cell*, 12(1), pp.63–74.
Yang, S.-H., Sharrocks, A.D. & Whitmarsh, A.J., 2013. MAP kinase signalling cascades and transcriptional regulation. *Gene*, 513(1), pp.1–13.
Yang, Z. et al., 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Molecular Cell*, 19(4), pp.535–545.
Yip, K.Y. et al., 2012. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome biology*, 13(9), p.R48.

Yoh, S.M. et al., 2007. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes & Development*, 21(2), pp.160–174.

Yoh, S.M., Lucas, J.S. & Jones, K.A., 2008. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes & Development*, 22(24), pp.3422–3434.

Yu, F.-X. et al., 2012. Regulation of the Hippo-YAP pathway by G-protein-coupled receptor signaling. *CELL*, 150(4), pp.780–791.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, 25(21), pp.2227–2241.

Zaromytidou, A.-I., Miralles, F. & Treisman, R., 2006. MAL and ternary complex factor use different mechanisms to contact a common surface on the serum response factor DNA-binding domain. *Molecular and cellular biology*, 26(11), pp.4134–4148.

Zehorai, E. et al., 2010. The subcellular localization of MEK and ERK--a novel nuclear translocation signal (NTS) paves a way to the nucleus. *Molecular and cellular endocrinology*, 314(2), pp.213–220.

Zeitlinger, J. et al., 2007. RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature genetics*, 39(12), pp.1512–1516.

Zhang, D.W. et al., 2012a. Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. *The Journal of biological chemistry*, 287(11), pp.8541–8551.

Zhang, H.-M. et al., 2008a. Mitogen-induced recruitment of ERK and MSK to SRE promoter complexes by ternary complex factor Elk-1. *Nucleic Acids Research*, 36(8), pp.2594–2607.

Zhang, H.M. et al., 2012b. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Research*, 40, pp.D144–9.

Zhang, J. & Corden, J.L., 1991. Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *The Journal of biological chemistry*, 266(4), pp.2297–2302.

Zhang, Y. et al., 2008b. Model-based analysis of ChIP-Seq (MACS). *Genome biology*, 9(9), p.R137.

Zhao, J. et al., 1999. Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor. *Molecular and cellular biology*, 19(11), pp.7733–7740.

Zheng, Y. et al., 2005. Phosphorylation of RasGRP3 on threonine 133 provides a mechanistic link between PKC and Ras signaling systems in B cells. *Blood*, 105(9), pp.3648–3654.

Zhong, W. et al., 2013. Mesenchymal stem cell and chondrocyte fates in a multishear microdevice are regulated by Yes-associated protein. *Stem cells and development*, 22(14), pp.2083–2093.

Zhu, J. et al., 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *CELL*, 152(3), pp.642–654.

Ziegler, E.C. & Ghosh, S., 2005. Regulating inducible transcription through controlled localization. *Science Signaling*, 2005(284), pp.re6–re6.

Zinck, R. et al., 1993. c-fos transcriptional activation and repression correlate temporally with the phosphorylation status of TCF. *The EMBO Journal*, 12(6), pp.2377–2387.

Zippo, A. et al., 2009. Histone Crosstalk between H3S10ph and H4K16ac Generates a Histone Code that Mediates Transcription Elongation. *CELL*, 138(6), pp.1122–1136.

Zippo, A. et al., 2007. PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. *Nature cell biology*, 9(8), pp.932–944.